1

1        IN THE UNITED STATES DISTRICT COURT
              NORTHERN DISTRICT OF ILLINOIS
2                  EASTERN DIVISION

3

KLEEN PRODUCTS, LLC, et al.,        )  Docket No. 10 C 5711
4                                   )
                    Plaintiffs,     )
5                                   )
            vs.                     )
6                                   )
PACKAGING CORPORATION OF AMERICA,   )  Chicago, Illinois
7  et al.,                          )  March 28, 2012
                                    )  8:00 o'clock a.m.
8                   Defendants.     )

9
            TRANSCRIPT OF PROCEEDINGS - EVIDENTIARY HEARING
10   BEFORE THE HONORABLE MAGISTRATE JUDGE NAN R. NOLAN
                        VOLUME 2-A
11
APPEARANCES:
12

13  For the Plaintiffs:      THE MOGIN LAW FIRM
                             BY:  MR. DANIEL J. MOGIN
14                           707 Broadway, Suite 1000
                             San Diego, CA  92101
15                           (619) 687-6611

16                           FREED KANNER LONDON & MILLEN LLC
                             BY:  MR. MICHAEL J. FREED
17                           MR. ROBERT J. WOZNIAK
                             2201 Waukegan Road, Suite 130
18                           Bannockburn, IL  60015
                             (224) 632-4500
19
                             SCOTT & SCOTT
20                           BY:  MR. WALTER W. NOSS
                             707 Broadway, Suite 1000
21                           San Diego, CA  92101
                             (619) 233-4565
22

23
Court Reporter:      MS. CAROLYN R. COX, CSR, RPR, CRR, FCRR
24                   Official Court Reporter
                     219 S. Dearborn Street, Suite 1854-B
25                   Chicago, Illinois  60604
                     (312) 435-5639

```
 1   APPEARANCES CONTINUED:

 2   For the Plaintiffs:          LOCKRIDGE GRINDAL NAUEN PLLP
                                  BY:  MR. BRIAN D. CLARK
 3                                100 Washington Avenue S.
                                  Minneapolis, MN  55401
 4                                (612) 239-6900

 5
     For Defendant Packaging      KIRKLAND & ELLIS LLP
 6   Corporation of America:      BY:  MR. BARACK S. ECHOLS
                                       MR. LEONID FELLER
 7                                300 North LaSalle Street
                                  Chicago, IL  60654
 8                                (312) 862-3144

 9
     For Defendant               FOLEY & LARDNER LLP
10   International Paper:         BY:  MR. JAMES T. McKEOWN
                                       MR. NATHAN EIMER
11                                777 East Wisconsin Avenue
                                  Milwaukee, WI  53202
12                                (414) 297-5530

13
                                  FOLEY & LARDNER LLP
14                                BY:  MS. JOANNE LEE
                                  321 North Clark Street, Suite 2800
15                                Chicago, IL  53202

16   For Defendant               MAYER BROWN LLP
     Temple-Inland:              BY:  MR. ANDREW S. MAROVITZ
17                                     MS. BRITT M. MILLER
                                  71 South Wacker Drive
18                                Chicago, IL  60606
                                  (312) 782-0600

19
     For Defendant               K & L GATES LLP
20   Cascades and Norampac:      BY:  MR. SCOTT M. MENDEL
                                  70 West Madison Street, Suite 3100
21                                Chicago, IL  60602
                                  (312) 372-1121

22

23   For Defendant               QUINN EMANUEL URQUHART &
     Georgia-Pacific:            SULLIVAN LLP
24                                BY:  MR. STEPHEN R. NEUWIRTH
                                  51 Madison Avenue, 22nd Floor
25                                New York, NY  10010
                                  212-849-7000
```

1 | APPEARANCES CONTINUED:

2 | For Defendant WINSTON & STRAWN LLP
RockTenn CP, LLC: BY: MR. R. MARK McCARIENS
3 | MR. JOSEPH L. SIDERS
35 West Wacker Drive
4 | Chicago, IL 60601
(312) 558-5902

5 |

6 | For Defendant McDERMOTT WILL & EMERY LLP
Wayerhaeuser Company: BY: MS. RACHAEL V. LEWIS
7 | 600 13th Street NW
Washington, DC 2005
8 | (202) 756-8479

9 |

10 |

11 |

12 |

13 |

14 |

15 |

16 |

17 |

18 |

19 |

20 |

21 |

22 |

23 |

24 |

25 |

4

1                                   I N D E X

2

3

4    DESCRIPTION                                                      PAGE

5

17

18

19

20

21

22

23

24

25

| | | |
|---|---|---|
| 09:03:16 | 1 | (The following proceedings were had in open court:) |
| 09:03:16 | 2 | THE CLERK:  10 C 5711, Kleen Products, et al., v. |
| 09:03:20 | 3 | Packaging Corporation of America, et al. |
| 09:03:22 | 4 | THE COURT:  Okay.  Good morning.  I am Nan Nolan.  I |
| 09:03:28 | 5 | am a magistrate judge.  Judge Feinerman was kind enough to |
| 09:03:32 | 6 | lend us his courtroom here.  And this case was referred from |
| 09:03:40 | 7 | Judge Shadur for all discovery matters, and we are in our |
| 09:03:42 | 8 | second day of an evidentiary hearing regarding word search in |
| 09:03:48 | 9 | this case. |
| 09:03:48 | 10 | So for the plaintiffs, will the plaintiffs' lead |
| 09:03:54 | 11 | lawyer introduce himself and your team, please. |
| 09:03:58 | 12 | MR. MOGIN:  Good morning, your Honor; Dan Mogin on |
| 09:04:02 | 13 | behalf of plaintiffs.  With me today, going around the table, |
| 09:04:08 | 14 | Walter Noss from the firm of Scott & Scott; Robert Wozniak |
| 09:04:10 | 15 | from the Freed Kanner firm; and Brian Clark from the Lockridge |
| 09:04:14 | 16 | firm. |
| 09:04:14 | 17 | THE COURT:  Okay.  Now, do these gentlemen -- |
| 09:04:20 | 18 | gentlemen, I will ask you directly, do you represent other |
| 09:04:22 | 19 | named plaintiffs in the case?  I know Mr. Mogin was appointed |
| 09:04:30 | 20 | as lead counsel for the putative class.  What I am confused |
| 09:04:38 | 21 | from from the docket when I was studying the docket in between |
| 09:04:44 | 22 | are there appear there are a number of other lawyers who |
| 09:04:46 | 23 | represent other plaintiffs.  Are you fellows representing the |
| 09:04:52 | 24 | other plaintiffs, or are you working with Mr. Mogin on this? |
| 09:04:56 | 25 | MR. MOGIN:  They are all working with me, your Honor. |

09:04:58　　1　It's just that what you viewed in the docket I think is just a

09:05:00　　2　matter of the way the docket gets entered.  All the

09:05:04　　3　plaintiffs' lawyers appear on the consolidated amended

09:05:08　　4　complaint.

09:05:08　　5　　　　　　THE COURT:  All right.  So you have -- for purposes

09:05:10　　6　of this hearing, though, I want to clarify, you have authority

09:05:14　　7　from -- I know this is funny with lead plaintiff, but you have

09:05:18　　8　authority from each of the other plaintiffs' lawyers that

09:05:26　　9　you're speaking on their behalf too?

09:05:28　10　　　　　　MR. MOGIN:  Yes, your Honor.

09:05:28　11　　　　　　THE COURT:  Okay.  Good.  Thank you.

09:05:30　12　　　　　　So for our defendants, will you state your name, your

09:05:36　13　partners -- I mean whoever is here with you, and what

09:05:42　14　defendant you represent, please.

09:05:44　15　　　　　　MR. NEUWIRTH:  Good morning, your Honor; Steven

09:05:46　16　Neuwirth, representing defendant Georgia-Pacific, and I am

09:05:50　17　joined here today by my colleague, Sammy Rasheed (phonetic).

09:05:54　18　　　　　　THE COURT:  Thank you, Mr. Neuwirth.

09:05:56　19　　　　　　MR. ECHOLS:  Good morning, your Honor; Barack Echols,

09:06:00　20　representing defendant Packaging Corporation of America, and I

09:06:02　21　am joined by my colleague, Leonid Feller.

09:06:04　22　　　　　　THE COURT:  Thank you.

09:06:08　23　　　　　　MR. McCAREINS:  Good morning, your Honor; Mark

09:06:10　24　McCareins from Winston Strawn on behalf of RockTenn, and I am

09:06:14　25　here with my colleague, Joe Siders.

| | | |
|---|---|---|
| 09:06:16 | 1 | THE COURT: Good. Thank you. |
| 09:06:18 | 2 | MR. EIMER: Good morning, your Honor; Nate Eimer on |
| 09:06:20 | 3 | behalf of International Paper. |
| 09:06:20 | 4 | THE COURT: Hi, Mr. Eimer. |
| 09:06:24 | 5 | MR. McKEOWN: Good morning, your Honor; James |
| 09:06:24 | 6 | McKeown. I am here with my colleague, Joanne Lee, for |
| 09:06:28 | 7 | International Paper. |
| 09:06:30 | 8 | THE COURT: Thank you. |
| 09:06:30 | 9 | MR. MAROVITZ: Good morning, your Honor; Andy |
| 09:06:30 | 10 | Marovitz for Temple-Inland, and I am here with my partner, |
| 09:06:36 | 11 | Britt Miller. |
| 09:06:36 | 12 | THE COURT: Thank you. And, Ms. Miller, good |
| 09:06:38 | 13 | morning. |
| 09:05:58 | 14 | MR. MENDEL: Good morning, your Honor; Scott Mendel |
| 09:06:42 | 15 | for Cascades and Norampac. |
| 09:06:42 | 16 | THE COURT: Sir, can you spell your last name. |
| 09:06:48 | 17 | MR. MENDEL: M-e-n-d-e-l. |
| 09:06:50 | 18 | THE COURT: Thank you, Mr. Mendel. And your client |
| 09:06:50 | 19 | is? |
| 09:06:52 | 20 | MR. MENDEL: Cascades and Norampac. |
| 09:06:54 | 21 | THE COURT: Thank you. |
| 09:06:54 | 22 | MS. LEWIS: Good morning, your Honor; Rachel Lewis on |
| 09:06:58 | 23 | behalf of Weyerhaeuser Company. |
| 09:07:00 | 24 | THE COURT: Okay. And that's everyone we have. |
| 09:07:06 | 25 | Okay. Good. |

| | | |
|---|---|---|
| 09:07:08 | 1 | Well, I have a few preliminary matters I'd like to |
| 09:07:14 | 2 | discuss, and you don't have to be shy.  If you want to add |
| 09:07:22 | 3 | anything or if you have any questions, please do. |
| 09:07:26 | 4 | So the first thing is Mr. Mogin told us the last |
| 09:07:36 | 5 | time -- I want to talk about how we are going to work today. |
| 09:07:40 | 6 | When we were here the last time, Mr. Mogin told us that either |
| 09:07:44 | 7 | it was a Ninth Circuit or California or district court meeting |
| 09:07:48 | 8 | which he needed to be at.  Based upon that, I went and |
| 09:07:50 | 9 | scheduled things and I have a 4:00 o'clock that's unmovable, |
| 09:07:56 | 10 | and so we are going to work -- what I would like to do today |
| 09:08:02 | 11 | is work here in the courtroom until about 10 to 3:00, we will |
| 09:08:08 | 12 | take a little walk around-the-table break at 10 to 3:00, and |
| 09:08:14 | 13 | then I'd like to invite you down to my courtroom with the |
| 09:08:18 | 14 | court reporter, and I'd like to have a conference, a |
| 09:08:22 | 15 | scheduling conference, with the lawyers at 3:00 until |
| 09:08:28 | 16 | approximately 10 to 4:00.  Okay? |
| 09:08:32 | 17 | So we will do as much as we can witness-wise given |
| 09:08:38 | 18 | that confine. |
| 09:08:38 | 19 | My thinking on this, this morning we should begin |
| 09:08:46 | 20 | with Mr. Regard because it was -- we cut him off or I cut him |
| 09:08:52 | 21 | off or he was cut off in the -- at the beginning of his direct |
| 09:09:00 | 22 | examination, and if we began with Mr. Regard, then he could be |
| 09:09:04 | 23 | crossed.  Dr. Lewis has been listening to what his comments, |
| 09:09:10 | 24 | at least the last time, he can then hear what Mr. Regard says, |
| 09:09:14 | 25 | and then he can opine about that.  Mr. Regard will be here. |

09:09:18　1　If there's any follow-up that Mr. Regard needs to do, we

09:09:22　2　certainly will give Mr. Regard an opportunity to do that.

09:09:26　3　　　　　Okay.  I know Dr. Tenny is here too, and as I said to

09:09:32　4　you, I did not know whether we'd have sufficient time today,

09:09:36　5　nobody is going to be precluded if we need more evidence.  I

09:09:42　6　am sure she is just finding this so stimulating that it will

09:09:50　7　be fine.

09:09:50　8　　　　　MR. McKEOWN:  Your Honor, if I might, would it be

09:09:52　9　more efficient to do Dr. Lewis first?  We are in the middle of

09:09:56　10　his cross.  We didn't finish him last time.

09:09:56　11　　　　　THE COURT:  Well, then we are going to have -- then I

09:09:58　12　think -- I mean, we stopped both of them.

09:10:02　13　　　　　MR. McKEOWN:  Correct.  We had Dr. Lewis' direct was

09:10:04　14　done.  We were in the middle of his cross.  And we at least

09:10:08　15　had anticipated that what would happen is we would finish

09:10:10　16　Dr. Lewis' cross and go back to Mr. Regard.

09:10:14　17　　　　　THE COURT:  You don't want anything Mr. Regard says

09:10:18　18　first?

09:10:20　19　　　　　MR. McKEOWN:  I don't think it's needed at this

09:10:20　20　point, from our perspective.

09:10:24　21　　　　　MR. MOGIN:  Your Honor, I think you're correct.  It

09:10:26　22　would be more efficient to have Mr. Regard go first.  But we

09:10:30　23　are willing to be flexible on that.

09:10:34　24　　　　　THE COURT:  How many people are crossing Dr. Lewis?

09:10:38　25　　　　　MR. McKEOWN:  I have to finish my cross, and then

09:10:40   1   it's just Mr. Neuwirth with the Georgia-Pacific's aspect.

09:10:44   2           THE COURT: Well, then maybe we should do it that

09:10:46   3   way. Maybe we have a better sense of timing then. Let's do

09:10:50   4   that. Okay?

09:10:52   5           And if -- so if we do Dr. Lewis, then we're going to

09:10:56   6   do Mr. Regard, then if Dr. Lewis wants to come back on and

09:11:00   7   comment on anything Mr. Regard says, if he wants to add that

09:11:06   8   to his opinion, we will allow him to do that. Okay?

09:11:08   9           MR. MOGIN: Very good. Thank you, your Honor.

09:11:10   10           THE COURT: That will be fine.

09:11:10   11           MR. MOGIN: Your Honor, before we do proceed with

09:11:12   12   Mr. Regard, there are several issues with respect to the

09:11:16   13   contours of his testimony we should take up.

09:11:18   14           THE COURT: With regard to? I'm sorry.

09:11:20   15           MR. MOGIN: Well, for example, we had made a motion

09:11:24   16   in limine.

09:11:24   17           THE COURT: Right.

09:11:24   18           MR. MOGIN: There are some aspects -- I am a little

09:11:28   19   confused at this point because Mr. Regard has put in some

09:11:32   20   direct testimony --

09:11:34   21           THE COURT: Right.

09:11:34   22           MR. MOGIN: -- that subsequently there was a report

09:11:36   23   which --

09:11:36   24           THE COURT: Right, which you object to.

09:11:40   25           MR. MOGIN: I object to -- I certainly object to the

09:11:42   1   chart.

09:11:42   2           THE COURT:  Right.

09:11:42   3           MR. MOGIN:  And I object to other aspects of the

09:11:44   4   report.

09:11:46   5           THE COURT:  Right.

09:11:46   6           MR. MOGIN:  So I just need -- it seems to me it's a

09:11:50   7   bit of a squishy target and I just want to know what the

09:11:52   8   ground rules are.

09:11:54   9           THE COURT:  I am going to -- I was waiting on that

09:11:56  10   one.  I was waiting to be -- I have a couple more housekeeping

09:12:00  11   things, and then I will get to that.  Okay?

09:12:04  12          So I wanted to put on the record, just like I put on

09:12:14  13   the record before, that I have known Mr. Regard for a while.

09:12:20  14   Last week I was lucky enough to go to a Sedona Conference, and

09:12:26  15   Dr. Lewis was one of the participants, not a speaker, but one

09:12:32  16   of the participants at the conference, so I had a nice

09:12:36  17   opportunity to have a few chatty words with Dr. Lewis.

09:12:42  18   Obviously, we did not talk about the case, but I did want to

09:12:44  19   tell you that.  Okay?

09:12:54  20          The exhibits, I had to actually find out when I was

09:12:56  21   on your side of the podium, the court reporters used to take

09:13:00  22   the exhibits.  And I know from the trials that I just want to

09:13:04  23   make absolutely sure you are responsible for your own

09:13:06  24   exhibits.  Okay?

09:13:10  25          So the nice books that the two of you prepared, as

| 09:13:16 | 1 | long as they are joint books, they will be fine for the last |

09:13:16   1   long as they are joint books, they will be fine for the last

09:13:20   2   hearing.  Thank you, Ms. Miller, for arranging all that for

09:13:26   3   us.  As you probably know, the clerk's office can't take bound

09:13:30   4   copies, so what I would do is -- what I think we will do is we

09:13:36   5   will have those -- if this issue ever goes to either Judge

09:13:42   6   Shadur or another court, you've got last time's, today we will

09:13:48   7   keep the original all together, one of you will be the

09:13:52   8   custodian of the original exhibits and maybe you will be kind

09:14:02   9   enough, if we need it, somebody will make us a copy too.

09:14:06   10   Thank you, Ms. Miller.  She is volunteering again.

09:14:08   11   Okay.  Let's see.  So last time, I know that I am

09:14:18   12   treating this much looser than I would treat a trial on ruling

09:14:24   13   on objections immediately.  I am kind of really -- it's very

09:14:32   14   important to me with all these folks from out of town that we

09:14:36   15   use every single minute that we have here doing the hearing,

09:14:40   16   so I was -- I know there were a number of objections, some of

09:14:46   17   the exhibits I did not rule on, some of that was wait until

09:14:52   18   the cross is over, so I will make any very best effort.  Chris

09:14:58   19   went through the transcript yesterday because it's searchable,

09:15:02   20   and we will try to clear it up.  And if I forget anything,

09:15:08   21   please feel free to let me know.

09:15:10   22   Okay.  So, Mr. Mogin, I received your motion in

09:15:18   23   limine regarding Mr. Regard.  I received the defendant's

09:15:24   24   response yesterday.  I assume that you did too.  So I am

09:15:28   25   denying your motion without prejudice.  Mr. Regard is -- so I

09:15:38    1    am going to allow him to testify today.

09:15:42    2         And on the chart, I am not ruling that I am admitting

09:15:48    3    the chart.  But the defendants even said -- I thought that was

09:15:54    4    a good reply -- I thought that was a good reply.  I mean, they

09:15:58    5    almost argued that it could be a demonstrative exhibit, and

09:16:04    6    demonstrative exhibits, I don't know how we are going sort all

09:16:08    7    this out.  I understand what you're saying.  I am sure you are

09:16:14    8    going to cross-examine him about many of the issues that you

09:16:16    9    raised in your motion.  I don't know if they go to weight, I

09:16:22   10    don't know whether it goes to admissibility, I haven't had

09:16:26   11    enough time to really sort that out, and we will, if we need

09:16:32   12    any follow-up after we hear what he's got to say today,

09:16:36   13    because I think almost any judge in this building would say,

09:16:40   14    How can I -- I am not going to preclude Mr. Regard up front,

09:16:46   15    and the exhibit, now that we have it and we know what goes

09:16:52   16    with it, I think I will be in a better position after

09:16:58   17    cross-examination to know what to do with it.

09:17:00   18         MR. MOGIN:  I understand your position, your Honor,

09:17:02   19    and, obviously, it will be not efficient for me to argue the

09:17:08   20    motion at this point in time.  However, I will say that the

09:17:10   21    issue is not the information that we do have.

09:17:12   22         THE COURT:  Right.

09:17:14   23         MR. MOGIN:  It's the information that we don't have,

09:17:16   24    which has some bearing on my ability to cross-examine

09:17:22   25    Mr. Regard.

| | | |
|---|---|---|
| 09:17:22 | 1 | So if Mr. Regard is going to testify about the bottom |
| 09:17:24 | 2 | part of the chart, I still do not have any documentation in |
| 09:17:28 | 3 | support of that. I have merely the hearsay. That's all we |
| 09:17:32 | 4 | have. |
| 09:17:34 | 5 | MR. MAROVITZ: Your Honor, Andy Marovitz for |
| 09:17:38 | 6 | Temple-Inland. Mr. Mogin has all of the documents that |
| 09:17:42 | 7 | Mr. Regard has seen in this connection. There's not been |
| 09:17:46 | 8 | anything that was provided to Mr. Regard in connection with |
| 09:17:48 | 9 | the preparation of the chart that Mr. Mogin does not have. |
| 09:17:54 | 10 | MR. MOGIN: Well, in point in fact, your Honor, |
| 09:17:56 | 11 | Mr. Regard testified before that he received documents, he |
| 09:18:02 | 12 | compiled his charts from documents, information that he |
| 09:18:04 | 13 | received from vendors that has not been produced. |
| 09:18:08 | 14 | THE COURT: Well, I think that's what you are going |
| 09:18:10 | 15 | to flush out. He provided -- one of the things I found to be |
| 09:18:14 | 16 | very helpful is he actually listed every single thing he read |
| 09:18:22 | 17 | in preparation. Now, did he have conversations -- I don't |
| 09:18:26 | 18 | know. That's what we are going to find out. Okay? And then |
| 09:18:30 | 19 | if he did not -- he has a page and a half single space of |
| 09:18:36 | 20 | every single thing that he read. But I am hearing you. You |
| 09:18:44 | 21 | know, there's a wonderful ability of a judge to strike things. |
| 09:18:50 | 22 | Okay? And we don't have a jury, and this is actually -- I am |
| 09:18:56 | 23 | so glad we have this hearing. That's what I am going to say |
| 09:18:58 | 24 | to you later on. I feel like I know the case so much better, |
| 09:19:00 | 25 | and it's been a real education for me. And you don't know me |

| | | |
|---|---|---|
| 09:19:06 | 1 | because this is your first case with me, but I am a pretty |
| 09:19:10 | 2 | fair person.  I just think we should get our witnesses on |
| 09:19:16 | 3 | today and get going and we will figure out what to do about it |
| 09:19:20 | 4 | in fairness afterwards.  Okay? |
| 09:19:24 | 5 | MR. MOGIN:  Thank you. |
| 09:19:24 | 6 | THE COURT:  Anybody else have anything you'd like to |
| 09:19:28 | 7 | say? |
| 09:19:30 | 8 | And any issues -- this is the same thing, Mr. Mogin. |
| 09:19:38 | 9 | If I reserved any issues the last time too, we will deal with |
| 09:19:42 | 10 | those at the end of the day too, unless there's something that |
| 09:19:46 | 11 | anybody is remembering right off the top of their head.  Okay? |
| 09:19:52 | 12 | All right.  So who did we say is going first?  Oh, |
| 09:19:56 | 13 | Dr. Lewis is coming back.  Okay. |
| 09:19:58 | 14 | Good morning, Doctor.  Come on up.  Lynette is not |
| 09:20:02 | 15 | here. |
| 09:20:34 | 16 | (Witness sworn.) |
| 09:20:34 | 17 | - - - |
| 09:20:34 | 18 | DANIEL D. LEWIS, CROSS-EXAMINATION CONTINUED |
| 09:20:34 | 19 | BY MR. McKEOWN: |
| 09:20:50 | 20 | Q.  Good morning, Dr. Lewis. |
| 09:20:50 | 21 | A.  Good morning, Counsel. |
| 09:20:52 | 22 | Q.  We are going to spend some time this morning on |
| 09:20:54 | 23 | Plaintiffs' Exhibit 10, which was the plaintiffs' proposed |
| 09:20:58 | 24 | CBAA search process.  Do you have a copy up there? |
| 09:21:04 | 25 | A.  I do not. |

09:21:06  1      MR. McKEOWN:  Your Honor, may I give the witness a

09:21:08  2  copy of the exhibit?

09:21:08  3      THE COURT:  Sure.  Of course.

09:21:10  4      THE WITNESS:  Thank you.

09:21:10  5      MR. McKEOWN:  Your Honor, did you need a copy of this

09:21:20  6  as well?

09:21:20  7      THE COURT:  Yes.  Thank you.

09:21:30  8  BY MR. McKEOWN:

09:21:30  9  Q.  Dr. Lewis, before we go through this process in a little

09:21:32  10  more detail, I would like to spend a little bit of time just

09:21:34  11  making sure that we are understanding the terminology in the

09:21:36  12  same way.  As you may recall, when we were last here in

09:21:38  13  February, we finished the day, we were talking about recall,

09:21:42  14  correct?

09:21:42  15  A.  Yes.

09:21:44  16  Q.  Okay.  And recall was of the group of documents that would

09:21:50  17  be responsive, what percentage are actually captured by the

09:21:54  18  information retrieval technology; is that fair?

09:21:56  19  A.  Yes, if one is talking about the recall and the

09:22:02  20  technology, it would be the set of documents that are in the

09:22:04  21  retrieve set or the retrieve set for that technology.

09:22:06  22  Q.  And we also talked about precision, and precision was the

09:22:10  23  concept of -- of the documents that are identified by our

09:22:14  24  search methodology, what percentage are actually responsive of

09:22:18  25  that set, correct?

09:22:20  1   A.  Right.  So within the set, again, we are looking at the

09:22:24  2   responsive documents within the retrieve set, and in this

09:22:26  3   case, we are comparing it to the size of that set rather than

09:22:30  4   in the case of recall, we compare it to all of the responsive

09:22:34  5   documents in the collection.

09:22:34  6   Q.  Now, you also used a term in your direct testimony called

09:22:40  7   a "confidence interval."  Do I understand correctly that a

09:22:44  8   confidence interval is the expected range within which you

09:22:48  9   expect the actual value to fall?

09:22:50  10  A.  It's a range in which, with high probability over a large

09:23:00  11  collection of random samples, you would expect the true value

09:23:04  12  to fall.

09:23:04  13  Q.  So, for example, if one were to take a political poll and

09:23:14  14  find that there was a sample that a particular candidate was

09:23:16  15  supported by 39 percent of the population, plus or minus 3

09:23:20  16  percent, let's just take that as our hypothetical.  Okay?  If

09:23:24  17  it's plus or minus 3 percent and the number is 39, the -- that

09:23:34  18  point estimate would be -- the central value would be 39,

09:23:40  19  correct?

09:23:40  20  A.  Yes, in your example, the central value is 39 percent.

09:23:44  21  Q.  And the confidence interval would give me the plus or

09:23:48  22  minus, correct, so it would go up to 42 or down to 36?

09:23:52  23  A.  That's correct.

09:23:52  24  Q.  And so that if I wanted to know the entire confidence

09:23:56  25  interval, I would say it's from 36 to 42?

09:23:58  1   A.  That's correct.

09:24:00  2   Q.  Now, there's also something that I saw that you referred

09:24:06  3   to as the margin of error.  Is that 3 percent in my example,

09:24:10  4   would that be called the margin of error?

09:24:12  5   A.  Yes.  That's -- essentially, the margin of error is -- the

09:24:18  6   usual interpretation of that is half the width of the

09:24:22  7   confidence interval.

09:24:22  8   Q.  Now, there's also a term you used called the confidence

09:24:28  9   level; is that correct?  Is that a statistical term?

09:24:32  10  A.  Yes, it is.

09:24:32  11  Q.  And I believe that in Plaintiffs' Exhibit 10, the proposed

09:24:38  12  protocol, you had proposed a confidence level of 95 percent;

09:24:38  13  is that right?

09:24:44  14  A.  I believe that there are two different confidence levels

09:24:50  15  in the protocol, so could you point me to the place where you

09:24:52  16  are talking about?

09:24:54  17  Q.  Sure.  I was looking at the page 2, step 7.

09:25:06  18  A.  Yes.

09:25:06  19  Q.  Where it says, A 95 percent confidence level?

09:25:10  20  A.  Yes.

09:25:10  21  Q.  For purposes of when we talk about a 95 percent confidence

09:25:14  22  interval, or a confidence level of 95 percent, am I correct

09:25:18  23  that what that means is that as a matter of statistics, you

09:25:24  24  have used a large enough sample so that you can be 95 percent

09:25:28  25  confident that the actual value falls within that confidence

09:25:32  1  interval?

09:25:32  2  A.  No.

09:25:34  3  Q.  Okay.  What's wrong with that description?

09:25:38  4  A.  You're describing it, the 95 percent, as if it's the

09:25:44  5  probability that the true value falls within the confidence

09:25:48  6  interval.  The 95 percent confidence level refers to over a

09:25:54  7  large collection of random samples, 95 percent of the random

09:26:00  8  samples would produce a confidence interval that contains the

09:26:04  9  true value.

09:26:04  10  Q.  Would the confidence level also indicate to you whether

09:26:10  11  the size of the sample or the sample was representative of the

09:26:12  12  total universe you have talked from?

09:26:14  13  A.  The appropriately drawn random sample of a certain size

09:26:24  14  will give you a confidence level.  Obviously, the process by

09:26:28  15  which the sample is selected has to be correct, the population

09:26:32  16  that it's selected from has to be appropriate.

09:26:34  17  Q.  Okay.  So that the 95 percent confidence level, as I

09:26:38  18  understand your testimony, is that if you kept drawing random

09:26:42  19  samples from this universe, I kept going out and polling the

09:26:46  20  population as to how many people support this particular

09:26:48  21  candidate, 95 percent of the time, the average I get from that

09:26:54  22  or the results I get from that sample are going to fall in my

09:26:56  23  hypothetical between 36 and 42 percent; is that right?

09:27:02  24  A.  No.

09:27:02  25  Q.  Okay.  I thought you said that the confidence level told

| | | |
|---|---|---|
| 09:27:08 | 1 | you -- let me back up. |
| 09:27:10 | 2 | I want to make sure we all understand.  What does it |
| 09:27:14 | 3 | mean when you say you have a confidence level of 95 percent? |
| 09:27:16 | 4 | A.  The confidence level is describing what the result of |
| 09:27:22 | 5 | repeated random samplings would be, and each random sample |
| 09:27:26 | 6 | would give you a different confidence interval.  And when you |
| 09:27:30 | 7 | say you have a 95 percent confidence interval, that means |
| 09:27:36 | 8 | you've used a procedure that 95 percent of the time would |
| 09:27:40 | 9 | produce a confidence interval that contains the correct value. |
| 09:27:46 | 10 | The -- |
| 09:27:46 | 11 | Q.  When you say -- let me interrupt you. |
| 09:27:48 | 12 | A.  Yes. |
| 09:27:48 | 13 | Q.  When you say "correct value," are you talking about the |
| 09:27:50 | 14 | correct value from the total population? |
| 09:27:52 | 15 | A.  Yes, that it would contain -- 95 percent of the time, you |
| 09:27:58 | 16 | would have a confidence interval that would contain the |
| 09:28:00 | 17 | population value of the statistic. |
| 09:28:04 | 18 | Q.  I'm sorry.  I thought that's where I started.  Maybe I |
| 09:28:06 | 19 | didn't explain it very artfully. |
| 09:28:08 | 20 | So that if I do this right and I draw my random |
| 09:28:12 | 21 | sample and I calculate my confidence interval, 95 percent of |
| 09:28:16 | 22 | the time, I am going to have a confidence interval that will |
| 09:28:18 | 23 | include the actual value if I were to go out and talk to the |
| 09:28:22 | 24 | total population about which candidate they -- |
| 09:28:26 | 25 | A.  Yes, that's correct. |

09:28:26    1   Q. I'm sorry.

09:28:30    2       If I had a 99 percent confidence level, that would

09:28:34    3   mean that 99 percent of the time, if I went out and polled the

09:28:38    4   entire population, not just a random sample, that my sample

09:28:42    5   would give me that value within 95 percent of the time, right?

09:28:46    6   A. Yes. And, indeed, the confidence level sometimes is a

09:28:54    7   little informal in that it could be more than 99 percent of

09:28:58    8   the time. 99 percent is kind of the guarantee you have.

09:29:00    9   Q. Okay. Now, in order to determine whether you are going to

09:29:06   10   be 95 percent confident or 99 percent confident or some other

09:29:10   11   percentage confident, you need to have the appropriate size

09:29:14   12   random sample, correct?

09:29:16   13   A. That is correct. You need to have an appropriately sized

09:29:20   14   random sample that's been correctly selected.

09:29:22   15   Q. So you use a software, and I think you used a program

09:29:24   16   called Raosoft or you referenced a program named Raosoft in

09:29:28   17   your materials?

09:29:28   18   A. I referenced a program from a website called Raosoft that

09:29:34   19   I used to double-check my calculations. That's not a program

09:29:38   20   which selects random samples; that's a web-based program which

09:29:44   21   computes confidence intervals.

09:29:46   22   Q. But it tells you how large a random sample you would want?

09:29:48   23   A. You could use -- I believe there is an option to use that

09:29:54   24   for choosing how large the random sample is also.

09:29:56   25   Q. But that's not the only program you could use to determine

| | | |
|---|---|---|
| 09:30:00 | 1 | how big a sample you need in order to determine if you have a |
| 09:30:02 | 2 | random sample? |
| 09:30:04 | 3 | A.  No, there are many such programs, or one could do the |
| 09:30:10 | 4 | calculation by hand. |
| 09:30:12 | 5 | Q.  Now, Plaintiffs' Exhibit 10, the proposed protocol -- by |
| 09:30:16 | 6 | the way, have you done any work on this matter since we were |
| 09:30:20 | 7 | last here in February? |
| 09:30:20 | 8 | A.  Could you be more specific? |
| 09:30:24 | 9 | Q.  Sure.  Have you done any further research or work on this |
| 09:30:28 | 10 | project for this litigation since we were last here and you |
| 09:30:30 | 11 | testified last time? |
| 09:30:32 | 12 | MR. MOGIN:  Your Honor, I don't know precisely where |
| 09:30:34 | 13 | Mr. McKeown is going with this, but there is an expert witness |
| 09:30:38 | 14 | stipulation in this case that basically precludes the |
| 09:30:42 | 15 | discovery of expert work product. |
| 09:30:48 | 16 | MR. McKEOWN:  I am not asking for anything that would |
| 09:30:50 | 17 | be work product. |
| 09:30:56 | 18 | BY MR. McKEOWN: |
| 09:30:56 | 19 | Q.  Let me ask this.  Have you read your testimony from last |
| 09:30:58 | 20 | time? |
| 09:30:58 | 21 | A.  I did review the transcript, yes. |
| 09:31:00 | 22 | Q.  Have you done any other work on which you are going to |
| 09:31:04 | 23 | rely your opinions here today? |
| 09:31:06 | 24 | A.  That's somewhat difficult to answer.  I read literature |
| 09:31:18 | 25 | about information retrieval all the time, I read statistical |

09:31:18   1   literature.  I don't recall anything that has, you know, been

09:31:22   2   particularly informative, but, you know, I am always learning

09:31:26   3   about statistics and information retrieval, so it's a little

09:31:28   4   hard to answer that question.

09:31:30   5   Q.  Okay.

09:31:32   6        THE COURT:  Mr. Mogin, you're correct.  We do have

09:31:34   7   that stipulation.  And I think Dr. Lewis has answered this.

09:31:42   8   I'm assuming he's read Mr. Regard's report, but maybe I am

09:31:48   9   wrong.  I mean, that would be something new.  He sat through

09:31:52   10   two days at Sedona and listened to tons of people yammering on

09:31:58   11   about predictive coding.  You know, I don't know.  I mean, I

09:32:04   12   don't know what you -- that's a pretty broad question,

09:32:06   13   Mr. McKeown, what you're asking what he would have done

09:32:10   14   between now and then.

09:32:12   15        MR. McKEOWN:  I just want to know if he had done any

09:32:14   16   other work specific to this case that he was going to rely on

09:32:18   17   for opinions.

09:32:18   18        THE COURT:  Oh, that he is relying on for opinions?

09:32:22   19   Oh, that's different.  Okay.

09:32:24   20        THE WITNESS:  So if you could ask me about particular

09:32:26   21   documents, I'd be happy to answer.

09:32:28   22   BY MR. McKEOWN:

09:32:28   23   Q.  Well, have you read Mr. Regard's report?

09:32:32   24   A.  Yes, I have.

09:32:32   25   Q.  Okay.  Other than that, are there any other documents that

09:32:36  1  you're relying on in your opinions today, anything new?

09:32:42  2  A. Well, again, there's a lot of documents in the case. I

09:32:44  3  looked some more at some of these spreadsheets studying stuff.

09:32:50  4  There was a March 21st letter from Georgia-Pacific that I read

09:32:54  5  a couple of times. You know, I don't know that I remember the

09:32:58  6  exact list of all the things I read.

09:33:00  7  Q. Let's turn to Plaintiffs' Exhibit 10. And the first two

09:33:06  8  pages are laid out in terms of steps; is that correct?

09:33:10  9  There's a number of steps to this proposed process?

09:33:12  10  A. Yes, there are.

09:33:12  11  Q. Okay. And the first step is labeled Collection. And I

09:33:18  12  want to come back to collection in just a minute. But if we

09:33:20  13  look at step 2, test set creation, do you see that on the

09:33:26  14  exhibit?

09:33:26  15  A. Yes, I do.

09:33:28  16  Q. Now, you have in step 2 two methods for test set creation.

09:33:36  17  One is 2-A, the indirect method, and 2-B is the direct method;

09:33:40  18  is that correct?

09:33:40  19  A. Yes.

09:33:40  20  Q. I'd like to spend just a minute on the indirect method,

09:33:44  21  item 2-A.

09:33:44  22  A. Okay.

09:33:44  23  Q. Now, as I understand step 2-A, you're going to create a

09:33:52  24  test set by drawing a random sample, and you're going to draw

09:33:56  25  that from the collection of documents that have been

09:33:58    1   assembled; is that correct?

09:34:00    2   A.  That's what that step describes.

09:34:02    3   Q.  And it says, A random sample of 2,401 documents.

09:34:08    4       Is that number drawn to obtain a 95 percent

09:34:12    5   confidence level with a plus or minus 2 percent?

09:34:16    6   A.  Let's look at the note C-1.

09:34:26    7       Right.

09:34:26    8   Q.  Just for the record, C-1, although there are no page

09:34:30    9   numbers on this document, is on the fourth page, notes on

09:34:36   10   indirect method; is that right?

09:34:40   11   A.  Yes.  Yes.

09:34:40   12       So the size 2401 was determined by saying we want a

09:34:48   13   95 percent confidence interval on the proportion of responsive

09:34:54   14   documents in the collection, and we want a margin of error

09:34:58   15   of .02 on that confidence interval on the proportion of

09:35:02   16   responsive documents in the collection.

09:35:04   17   Q.  And so in my example before with the political survey, we

09:35:08   18   had a plus or minus 3 percent.  This would call for a plus or

09:35:12   19   minus 2 percent?

09:35:14   20   A.  That's correct.

09:35:14   21   Q.  And it would have the 95 percent confidence level we

09:35:18   22   talked about that you described for me before; is that

09:35:18   23   correct?

09:35:20   24   A.  That's correct.

09:35:20   25   Q.  Now, the purpose of this taking of this sample is in order

09:35:24　1　to give an estimate of what the total number of responsive

09:35:28　2　documents are in the entire collection; is that right?

09:35:32　3　A.　Yes, one can do it either as estimating total number

09:35:36　4　responsive or the proportion of responsive, but the two are

09:35:40　5　equivalent.

09:35:40　6　Q.　And so that what you're attempting here is you're going to

09:35:44　7　take this random sample, you're going to find out some number,

09:35:48　8　maybe it's 20 percent, maybe it's 15 percent, maybe it's 30

09:35:52　9　percent, and this sample is going to tell you that if you

09:35:54　10　reviewed all the documents in the entire universe of this

09:35:58　11　collection, that the true value would be 95 percent confident

09:36:02　12　falls somewhere, plus or minus 2 percent, of what you got from

09:36:06　13　that sample; is that correct?

09:36:08　14　A.　Yes.　So 95 percent of the time the random sample would

09:36:12　15　give you a confidence interval that contained the true value

09:36:14　16　of that.

09:36:14　17　Q.　And that's because you want to know how many -- what's the

09:36:18　18　total universe of documents, responsive documents, try and

09:36:22　19　make an estimate of that, right?

09:36:24　20　A.　Try to estimate in the universe of all the documents what

09:36:26　21　is the proportion of them which are responsive.

09:36:28　22　Q.　Now, you could do the same thing with a null set, couldn't

09:36:34　23　you?

09:36:34　24　A.　So first in my understanding is in this case, the term

09:36:50　25　"null set" is not being used in the set -- theoretic set of

09:36:54  1  the set that contains nothing.  It is being used to describe

09:36:58  2  the set-upon retrieved documents.

09:37:00  3  Q.  Fair.

09:37:00  4  A.  So the -- one potentially could sample only from the null

09:37:06  5  set.  However, there is an even stronger necessity in that

09:37:14  6  case of the consistency of the review of the sample and the

09:37:18  7  review for production, and those two reviews have to be

09:37:26  8  consistent or the estimate would be invalid.

09:37:28  9  Q.  Okay.  And let me -- let's try to take this a step at a

09:37:32  10 time.

09:37:32  11       So -- and I agree with your refinement of the null

09:37:36  12 set.  We are not talking about a set that has nothing.

09:37:38  13 A.  Right.

09:37:38  14 Q.  So let's say we take the documents, we have this

09:37:40  15 collection of documents, a number of them are hit by search

09:37:44  16 terms, those are sent off to the review process.

09:37:46  17       What we have left, this corpus, we are going to call

09:37:50  18 our null set, correct?  And just as you propose in the

09:37:56  19 indirect method in Plaintiffs' Exhibit 10 to draw a random

09:37:58  20 sample in order to estimate how many documents are in the

09:38:02  21 entire collection there, one could draw a random sample from

09:38:06  22 the null set, and assuming they were properly reviewed, you

09:38:10  23 would have a statistical estimate of how many responsive

09:38:16  24 documents remained in that null set, correct?

09:38:18  25 A.  Right.  So you would get a confidence interval on the

09:38:22  1  proportion and, thus, the number of the documents in the null

09:38:26  2  set -- the proportion of the number of responsive documents in

09:38:30  3  the null set.  And then if the review of that is consistent --

09:38:36  4  that sample is consistent with a review of the population,

09:38:40  5  and, that is, you're assuming you have a perfect review of the

09:38:44  6  production set, then you would have -- you could use that then

09:38:46  7  to have an estimate on the total number of responsive

09:38:50  8  documents in the collection.

09:38:52  9  Q.  Okay.  I'm not asking you to go to the next step,

09:38:56  10  Dr. Lewis.

09:38:56  11  A.  All right.

09:38:56  12  Q.  I'm not asking you to tell me what the total response is

09:39:00  13  in the whole universe.

09:39:00  14  A.  Just the null set.

09:39:02  15  Q.  I'm just looking at the null set.  As I look at the null

09:39:04  16  set, I draw a random sample, and I can find by that random

09:39:08  17  sample to a 95 percent confidence level in the example we have

09:39:10  18  just used where that 95 percent of the time, the confidence

09:39:16  19  level from my sample would include the true value in the null

09:39:18  20  set in the number of responsive documents that are left?

09:39:22  21  A.  Right.  So assuming a proper review, that's correct.

09:39:28  22  Q.  And the number you used for the indirect method in

09:39:32  23  Plaintiffs' Exhibit 10, the 2401 items in the random sample,

09:39:38  24  that was driven by the fact that you wanted a confidence level

09:39:40  25  of 95 percent and a margin of error, a plus or minus 2

09:39:46  1  percent, correct?

09:39:46  2  A.  That's correct.

09:39:48  3  Q.  I mean, there is a third factor; that would be what's the

09:39:50  4  size of the population you're sampling?

09:39:52  5  A.  Right.  The size of the population, given the sizes of

09:39:58  6  populations which are typically seen in e-discovery, would

09:40:02  7  only have a very small effect.  If you have a really small

09:40:04  8  population, sometimes you can shrink that sample down a

09:40:08  9  little.

09:40:08  10  Q.  So if, for example, instead of wanting to go 95 percent

09:40:12  11  confidence level with a plus or minus 2, you wanted to go to a

09:40:14  12  95 percent confidence level, plus or minus 5 percent, you

09:40:20  13  would need a sample size of only somewhere in the 370 to 385

09:40:24  14  range, correct?

09:40:24  15  A.  I would have to sit down and do the calculation to be

09:40:28  16  sure.  The exact size would be smaller.

09:40:30  17  Q.  If you wanted to do a 99 percent confidence level, plus or

09:40:36  18  minus 5 percent, would that require a sample size of

09:40:44  19  approximately 660 to 665?

09:40:46  20  A.  I would have to double-check that, but that sounds about

09:40:50  21  right.

09:40:50  22  Q.  And you know that in this case, the defendants drew random

09:41:00  23  samples from their null sets, correct?

09:41:02  24  A.  Yes, that's my understanding.

09:41:04  25  Q.  And you understood that the random samples were devised to

09:41:08  1   be of a size to give either 99 or 95 percent confidence level

09:41:12  2   with a plus or minus 5 margin of error, correct?

09:41:16  3           MR. MOGIN:  Objection, your Honor.  Counsel has used

09:41:20  4   the plural, and I am not sure if the plural refers to the

09:41:22  5   number of random samples per defendant or if it refers to the

09:41:28  6   fact that there are multiple defendants who each did a random

09:41:30  7   sample.

09:41:32  8           THE COURT:  All right.  So will you rephrase your

09:41:34  9   question?

09:41:36  10          MR. McKEOWN:  I will, your Honor.

09:41:36  11          THE COURT:  The objection is sustained.  Rephrase the

09:41:38  12  question.

09:41:38  13          MR. McKEOWN:  Yes, your Honor.

09:41:40  14  BY MR. McKEOWN:

09:41:42  15  Q.  You understand that at least some defendants used a sample

09:41:46  16  size that would give a 99 percent confidence level with a plus

09:41:50  17  or minus 5 percent margin of error on the null set, correct?

09:41:54  18  A.  I would like to see the document where you are describing

09:41:58  19  that.  There's been a number of documents that describe

09:42:00  20  different samples and confidence intervals and whatnot, so it

09:42:04  21  would probably be good to be sure that we were talking about

09:42:08  22  the same thing.

09:42:08  23  Q.  But -- and I want to move this along.  Theoretically --

09:42:12  24          THE COURT:  Now, wait, Mr. McKeown, because I am

09:42:14  25  following this.  We know Georgia-Pacific did that.  I mean, to

09:42:20  1    me, what we heard is Georgia-Pacific -- I thought that was

09:42:26  2    Georgia-Pacific's testimony.  I don't think we have anything

09:42:30  3    in from anybody else.

09:42:34  4           MR. McKEOWN:  Well, there have been a number of

09:42:36  5    letters that have been submitted by the defendants that

09:42:36  6    described their process, and this is the bottom of Mr.

09:42:40  7    Regard's chart.

09:42:40  8           THE COURT:  Okay.  So then --

09:42:46  9           MR. McKEOWN:  I can move on to the hypothetical, your

09:42:50  10   Honor.

09:42:50  11          THE COURT:  Well, I don't -- why don't we talk about

09:42:58  12   Georgia-Pacific?  I guess Mr. Neuwirth is going to talk about

09:42:58  13   Georgia-Pacific when he does it.

09:43:00  14          Now, do you want to show him each of the letters?  I

09:43:02  15   am not -- I mean, I don't think that's fair to a witness to

09:43:06  16   say, you know, Did X report a 95 percent -- I mean, I guess it

09:43:14  17   could be did X report a 95 percent confidence level or how

09:43:20  18   they did a random sample of their null set.

09:43:26  19          Did you want to say something?  I keep jumping.

09:43:28  20          MR. MOGIN:  No, your Honor.  I think you have hit

09:43:30  21   precisely on the problem.  Now we are into --

09:43:32  22          THE COURT:  Well, I don't know whether it's a problem

09:43:34  23   or whether it's just a matter here of being able to phrase it.

09:43:38  24   Okay?  This is where I think it could go to weight as opposed

09:43:44  25   to admissibility.  I mean, experts all the time -- Dr. Lewis

| | | |
|---|---|---|
| 09:43:48 | 1 | read a whole bunch of things here.  Okay? |
| 09:43:52 | 2 | So -- but I don't think it's fair unless you |
| 09:43:54 | 3 | specifically point him to a spot on the other defendants and |
| 09:44:00 | 4 | ask him if he's aware of that.  Maybe he did read your letter. |
| 09:44:06 | 5 | MR. McKEOWN:  Let me ask that question, your Honor. |
| 09:44:06 | 6 | THE COURT:  Okay.  You can sit down.  Thank you. |
| 09:44:12 | 7 | Because you are making me nervous standing up. |
| 09:44:14 | 8 | MR. MOGIN:  I don't want to make you nervous, your |
| 09:44:18 | 9 | Honor. |
| 09:44:18 | 10 | BY MR. McKEOWN: |
| 09:44:18 | 11 | Q.  Dr. Lewis, let me ask a couple questions. |
| 09:44:20 | 12 | A.  Sure. |
| 09:44:20 | 13 | Q.  Did you read the letters that the various defendants sent |
| 09:44:24 | 14 | in response to Ms. Vulgan's (phonetic) letters that were sent |
| 09:44:28 | 15 | last week by the various plaintiffs? |
| 09:44:30 | 16 | A.  I did at least skim over all of those letters.  I did not |
| 09:44:34 | 17 | read all of them in detail. |
| 09:44:38 | 18 | THE COURT:  So you're referring to -- okay.  That |
| 09:44:40 | 19 | helps me because I've got that somewhere.  So that question |
| 09:44:46 | 20 | regarding each of the defendant's random sample of null set is |
| 09:44:54 | 21 | in -- the answers came from Mr. Mogin's post hearing letter to |
| 09:45:02 | 22 | them. |
| 09:45:04 | 23 | MR. McKEOWN:  Part of the answer was there, yes, your |
| 09:45:06 | 24 | Honor. |
| 09:45:08 | 25 | THE COURT:  Okay.  I do have that someplace. |

09:45:22   1   BY MR. McKEOWN:

09:45:22   2   Q.  Dr. Lewis, have you read Mr. Regard's report?

09:45:24   3   A.  Yes, I have.

09:45:26   4   Q.  And have you read Mr. Regard's exhibits?

09:45:28   5   A.  Which exhibits are Mr. Regard's exhibits?

09:45:34   6   Q.  The documents that were attached to his report.

09:45:38   7   A.  The documents I recall being attached, there was like a

09:45:44   8   resume kind of thing and background information.  If you want,

09:45:50   9   I'd be happy to look and tell you whether I have looked at a

09:45:54  10   particular document.

09:45:54  11   Q.  Well, did you look at his chart, the large chart that

09:45:56  12   listed by every defendant how many documents were in the null

09:46:00  13   set, how many were sampled, what the results were of the null

09:46:04  14   set testing?

09:46:06  15   A.  Right.  So this was the chart that was introduced on

09:46:08  16   February -- the chart that was handed around the courtroom on

09:46:12  17   February 21st?

09:46:12  18   Q.  Correct.

09:46:14  19   A.  I did look at that piece of paper, yes.

09:46:16  20   Q.  In looking at that piece of paper, did you understand that

09:46:18  21   piece of paper, that chart, to show the sample size from the

09:46:24  22   null sets that each of the various defendants had drawn for

09:46:26  23   purposes of null set testimony?

09:46:30  24        MR. MOGIN:  Your Honor, at this point, I just want to

09:46:32  25   point out that Dr. Lewis is being asked to testify about

09:46:36   1    hearsay inside of hearsay inside of hearsay, and the

09:46:40   2    defendants are trying to get this in for the truth of the

09:46:44   3    matter in it; that is, that these were, in fact, the hit

09:46:50   4    counts, that these were, in fact, the tests that were

09:46:54   5    performed.  The letters themselves are hearsay within hearsay

09:46:58   6    and inadmissible.

09:47:16   7           MR. McKEOWN:  Your Honor, I don't think I was headed

09:47:18   8    where Mr. Mogin thinks I was headed.  I was about to wrap up

09:47:20   9    this line.  I just wanted to see if he had actually seen these

09:47:22  10    documents.

09:47:22  11           THE COURT:  Look, we can cut through -- I mean, we

09:47:26  12    can cut through some of this.  This is going to happen when

09:47:30  13    you have a summary witness.  Okay?

09:47:32  14           The alternative, if we decide after today that you

09:47:40  15    want each of the defendants to put a specific person on, this

09:47:44  16    is not that it's not curable here.  Okay?  I mean, I think

09:47:48  17    they could each put on a person as Georgia-Pacific did.

09:48:00  18           This entire issue has come in a different form than

09:48:06  19    most evidentiary issues.  I mean, this was not exactly the

09:48:14  20    normal procedural question that a court has to decide.  I

09:48:20  21    mean, as you know, it didn't even really come on a motion.  It

09:48:24  22    was a request from the defendants to see if the -- to test out

09:48:32  23    if their method was correct because in this case you were

09:48:36  24    suggesting a different method.

09:48:40  25           So I agree with you, Mr. Mogin, but I am still trying

09:48:44    1    to learn what has happened here.  Okay?

09:48:50    2            One alternative here could be to call all seven --

09:48:56    3    each of the seven defendants, put a person on who has personal

09:49:00    4    knowledge of what they did.  I don't know today if that's the

09:49:04    5    best use of our time.  So I am letting things come in

09:49:12    6    partially as an education for me of what the situation is here

09:49:16    7    because I wasn't in the case for the first nine months.  Now,

09:49:20    8    maybe someone might not think that's the appropriate role

09:49:28    9    here, but I think as we're going forward, that's very helpful

09:49:32    10   to the person who is, you know, charged with trying to manage

09:49:36    11   discovery here.

09:49:38    12           MR. MOGIN:  I understand, your Honor, and I

09:49:42    13   understand the value of the educational role.  But here in

09:49:46    14   this circumstance, recall what happened, it was the plaintiffs

09:49:52    15   who said, We are going to file a motion.  And the defendants

09:49:56    16   raced in and said, Oh, don't file a motion.  We want an

09:49:58    17   evidentiary hearing.  If they asked for the evidentiary

09:50:02    18   hearing, you can't have an evidentiary hearing without some

09:50:04    19   rules of evidence.

09:50:06    20           Now, I admit that the rules can be a little bit lax

09:50:10    21   under the circumstances.  But this is fundamental.  They are

09:50:14    22   trying to sneak in through the back door facts about the

09:50:18    23   testing, and then they are going to ask you to make a decision

09:50:22    24   based upon those facts.  We have no ability to cross-examine.

09:50:26    25   We don't have the documents, we don't have the percipient

09:50:30  1  witnesses.

09:50:30  2           THE COURT:  I just said to you that if it got to

09:50:36  3  that, if when I have to rule, the worst-case scenario here is

09:50:42  4  -- well, I suppose I could say, A, you didn't meet your

09:50:46  5  burden, or I could say, I'm opening up this hearing again and

09:50:48  6  they're calling all seven people if I need to do that.

09:50:52  7           Right now, this is helping me because I am struggling

09:50:58  8  with this very issue of random sample of null sets.  Okay?

09:51:04  9  Myself.

09:51:04  10          MR. MOGIN:  We too are struggling, your Honor.

09:51:06  11          THE COURT:  Fine.  Then we are on the same page.

09:51:10  12          So your objection right now is overruled for the

09:51:12  13  moment.  Okay.  Continue on.

09:51:14  14  BY MR. McKEOWN:

09:51:16  15  Q.  Dr. Lewis, we have spent more time on this than I had

09:51:18  16  intended, but my question was -- where I was headed was this.

09:51:24  17  If someone took a random sample of 660 or 665 documents out of

09:51:32  18  a null set of something greater than a hundred thousand

09:51:36  19  documents, they could produce a statistical result that would

09:51:44  20  give them a confidence interval, plus or minus 5 percent, with

09:51:48  21  a 99 percent confidence level; isn't that correct?

09:51:52  22  A.  You know, again, I would have to sit down and double-check

09:51:54  23  the calculation, but that sounds about the right size of the

09:51:58  24  sample you would need to get a confidence interval on the

09:52:02  25  proportion of responsive documents in the null set only.  It

09:52:06  1  obviously doesn't say anything about the hit set or about

09:52:08  2  recall.

09:52:08  3  Q.  Correct.  And I am not asking about the hit set or recall.

09:52:12  4  I am only asking about the null set.

09:52:14  5       So with that, you would have to do the calculation --

09:52:18  6  A.  I would have to do the calculation --

09:52:20  7  Q.  -- to see if it's 660 or 665?

09:52:22  8  A.  600 is on the order of the right number.

09:52:26  9  Q.  If we could turn back to Plaintiffs' Exhibit 10, the

09:52:28  10  protocol.  In step 1, the collection, it says, Determine

09:52:38  11  collection of documents to be searched for this defendant.

09:52:42  12       Now, I think you testified on direct that what you

09:52:46  13  assumed is that all reasonable sources of responsive documents

09:52:50  14  have been collected; is that right?

09:52:52  15  A.  I believe that's correct.

09:52:52  16  Q.  You haven't made any conclusions as to which documents are

09:52:58  17  to be collected; is that correct?

09:53:00  18  A.  That's correct.

09:53:02  19  Q.  You're offering no opinions as to which custodians should

09:53:04  20  be included or not included, correct?

09:53:06  21  A.  I believe the only thing that I have said about custodians

09:53:12  22  is that all of the custodians that are a reasonable source of

09:53:18  23  information should be sampled from it, the entire collection

09:53:22  24  should be sampled from if one was going to draw valid

09:53:26  25  conclusions.  I have not expressed opinions about any

09:53:28    1  particular custodians.

09:53:28    2  Q.  Or whether this particular custodian is a reasonable

09:53:34    3  custodian or is not a reasonable --

09:53:36    4  A.  No, I haven't been asked to look at that.

09:53:38    5  Q.  And you have not given any opinion as to what that number

09:53:42    6  is either, correct, in terms of how many custodians that

09:53:44    7  should be?

09:53:44    8  A.  That's correct.

09:53:46    9  Q.  And sitting here today, you don't have any opinion in

09:53:50   10  terms of what would be the total number of gigabytes or

09:53:54   11  terabytes that would be in the collection for each defendant

09:53:58   12  that's referenced in Plaintiffs' Exhibit 10 under step 1?

09:54:00   13  A.  That's correct.

09:54:00   14  Q.  You also offered no opinion, as I understood your direct

09:54:10   15  testimony, on what the cost to the defendants would be of

09:54:12   16  complying with what's been proposed in Plaintiffs' Exhibit 10;

09:54:16   17  is that correct?

09:54:16   18  A.  That's correct.

09:54:18   19  Q.  Now, in terms of collection, is it your understanding that

09:54:40   20  whatever the documents are, whatever the reasonable group of

09:54:44   21  custodians is, that those documents would then be taken and

09:54:46   22  put onto some vendor's system for purposes of this protocol?

09:54:52   23  A.  Well, there's a number of ways that you could do that.  I

09:54:56   24  mean, you could bring the software to the data in some sense,

09:55:02   25  or you could bring the data to the software, you could use an

| | | |
|---|---|---|
| 09:55:04 | 1 | outsourcing company or you could lease software.  In terms of |
| 09:55:08 | 2 | where the data physically resides, that can be done a lot of |
| 09:55:12 | 3 | different ways. |
| 09:55:12 | 4 | Q.  Okay.  But the data is going to need to be copied or |
| 09:55:14 | 5 | collected somehow to start this process, correct? |
| 09:55:16 | 6 | A.  What would be necessary is that the data be indexed and |
| 09:55:22 | 7 | made available to the tools that supports statistical rank |
| 09:55:30 | 8 | retrieval and supervised learning. |
| 09:55:30 | 9 | Q.  And that collection has to stay constant as well, correct? |
| 09:55:34 | 10 | A.  Well, from the standpoint of operating supervised learning |
| 09:55:44 | 11 | software or statistical rank retrieval software, most software |
| 09:55:50 | 12 | of that sort has the ability to bring new documents into the |
| 09:55:54 | 13 | collection. |
| 09:55:56 | 14 | From the standpoint of the statistical sampling, if |
| 09:56:02 | 15 | substantial numbers of the documents were brought in, the |
| 09:56:04 | 16 | sample would have to be either redone or augmented, and there |
| 09:56:10 | 17 | are statistical techniques that could be used so that you |
| 09:56:12 | 18 | might not have to redo the sampling from scratch.  But you do |
| 09:56:16 | 19 | need to make sure that you end up with a statistically |
| 09:56:22 | 20 | appropriate representative sample of the entire collection of |
| 09:56:26 | 21 | new documents that come in. |
| 09:56:28 | 22 | Q.  And maybe my question just wasn't very clear. |
| 09:56:30 | 23 | A.  Sorry. |
| 09:56:30 | 24 | Q.  Under Plaintiffs' Exhibit 10, this protocol, in the real |
| 09:56:34 | 25 | world, if one were to try to apply this, you have to have |

| | | |
|---|---|---|
| 09:56:38 | 1 | whatever the collection of documents is that you are going to |
| 09:56:40 | 2 | apply it to, correct? |
| 09:56:42 | 3 | A.  Right. |
| 09:56:42 | 4 | Q.  And that collection of documents is going to be copied and |
| 09:56:48 | 5 | it's not going to be changing during this protocol, correct? |
| 09:56:52 | 6 | A.  The collection of documents can change during the |
| 09:56:56 | 7 | protocol.  There's no difficulty with the operation of |
| 09:57:00 | 8 | statistical rank retrieval or supervised learning if the |
| 09:57:04 | 9 | document collection is changing.  There would be some |
| 09:57:06 | 10 | adjustments that would need to be made to the random sampling |
| 09:57:10 | 11 | in order to have a statistically valid estimate of the recall. |
| 09:57:18 | 12 | Q.  Is it your understanding that the documents would need to |
| 09:57:24 | 13 | be moved to some system, whether it's in-house or with a |
| 09:57:28 | 14 | vendor so that these processes could be applied? |
| 09:57:32 | 15 | A.  Not necessarily.  There are information retrieval systems |
| 09:57:42 | 16 | that are able to essentially sort of go out to where you point |
| 09:57:46 | 17 | them to on the network and index the documents in place. |
| 09:57:50 | 18 | Typically, the index itself would be built in a centralized |
| 09:57:58 | 19 | location, although there are some software that has |
| 09:58:00 | 20 | distributed indexing, but there's no necessity that the |
| 09:58:04 | 21 | documents physically would have to be moved. |
| 09:58:08 | 22 | Q.  Are you aware of that being done in the discovery context? |
| 09:58:10 | 23 | A.  To be -- I'm not sure if that's done.  I know that there |
| 09:58:18 | 24 | are enterprise knowledge management systems that allow search |
| 09:58:22 | 25 | over documents in place.  I am not sure if that's a capability |

| | | |
|---|---|---|
| 09:58:26 | 1 | that's in any of the current e-discovery software. |
| 09:58:28 | 2 | Q. Were you here during Mr. Hanners' testimony? |
| 09:58:30 | 3 | A. Yes, I was. |
| 09:58:32 | 4 | Q. And Mr. Hanners had some concern about collecting and |
| 09:58:34 | 5 | setting aside documents so that the metadata didn't change? |
| 09:58:38 | 6 | A. Right. I understand that there's a concern about opening |
| 09:58:44 | 7 | -- for instance, opening a document in some software |
| 09:58:46 | 8 | application that might change the metadata and wants to avoid |
| 09:58:50 | 9 | that. |
| 09:58:50 | 10 | Q. So whatever the universe is that you are going to apply |
| 09:58:54 | 11 | this protocol to, unless you somehow capture it and collect it |
| 09:58:56 | 12 | in a fixed snapshot, so to speak, you run that risk that |
| 09:59:02 | 13 | Mr. Hanners is concerned about, correct? |
| 09:59:04 | 14 | A. Well, again, that would depend on the details of the |
| 09:59:06 | 15 | operating system of the computer. There are ways to lock |
| 09:59:10 | 16 | files and make them read only so you would make them read only |
| 09:59:14 | 17 | so that they can't be changed. |
| 09:59:16 | 18 | So I am not -- I don't have an opinion about what |
| 09:59:18 | 19 | would be the most practical way to ensure that the metadata is |
| 09:59:24 | 20 | not changed. I am aware that there's several different ways |
| 09:59:26 | 21 | that that could be done, including copying the files. |
| 09:59:30 | 22 | Q. With any document collection in a normal litigation |
| 09:59:34 | 23 | context, there is usually cost of collection, correct? |
| 09:59:42 | 24 | A. I would assume that any activity costs something. |
| 09:59:44 | 25 | Q. When laptops are imaged, that costs money, correct? |

09:59:46  1  A.  Yes, I have certainly had that done, so I know that costs

09:59:50  2  money.

09:59:50  3  Q.  When servers are copied so that the documents can be sent

09:59:52  4  to a vendor for processing, that costs money?

09:59:56  5  A.  Very few things are free.  I agree to that.

10:00:00  6  Q.  And when documents go to an ESI vendor, they usually also

10:00:04  7  charge something for hosting your data; isn't that correct?

10:00:08  8  A.  That's my understanding.

10:00:08  9  Q.  Okay.  And that it can cost thousands or even more than

10:00:12  10  $10,000 a month just to host the data in a large case,

10:00:16  11  correct?

10:00:16  12  A.  I haven't looked into the particular figures.  I assume

10:00:20  13  they charge something.

10:00:20  14  Q.  And isn't it also the case that the various vendors who

10:00:26  15  offer this supervised learning charge an extra charge for

10:00:30  16  processing that data for the application of that software?

10:00:32  17  A.  I am not aware of the details of the charging for the

10:00:38  18  different options of the software services.

10:00:40  19  Q.  Now, you're a consultant for Kroll; is that correct?

10:00:44  20  A.  That's correct.

10:00:44  21  Q.  And are you familiar with Kroll's per gigabyte charge for

10:00:50  22  applying your software?

10:00:52  23  A.  I am not.

10:00:54  24  Q.  Are you aware of the charge by Recommind for applying

10:00:58  25  their software?

10:00:58  1  A.  No, I'm not.

10:01:00  2  Q.  Are you aware of the charge that OrcaTec charges for

10:01:04  3  applying a supervised learning software?

10:01:04  4  A.  No, I'm not.

10:01:06  5  Q.  Are you aware of the charge that any company has on per

10:01:10  6  gigabyte or some other basis for what it costs to apply this

10:01:14  7  supervised learning software to some collection of ESI?

10:01:18  8  A.  No, I'm not.

10:01:18  9  Q.  And sitting here today, you can't tell me if that's $400,

10:01:32  10  $600, or $900 a gigabyte to apply the supervised learning

10:01:40  11  software; is that correct?

10:01:42  12  A.  That's correct.  My understanding is that different

10:01:44  13  vendors indeed have different pricing schemes.  I don't know

10:01:46  14  the details of them, but some charge by volume, some charge by

10:01:50  15  disk space, some charge by numbers of documents, some charge

10:01:54  16  by a variety of other complicated formulas.

10:01:56  17  Q.  Whatever the charge is per gigabyte, you would agree that

10:02:08  18  the more gigabytes or more terabytes, the more expensive it's

10:02:12  19  going to be to apply, correct, the supervised learning

10:02:14  20  software?

10:02:16  21  A.  One would have to be careful about that in the sense that

10:02:20  22  different types of files take up vastly different amounts of

10:02:26  23  space.  So, for instance, if one had an archive of digital

10:02:30  24  video, that could be an extraordinarily large amount of space

10:02:34  25  but a rather small number of documents.

| | | |
|---|---|---|
| 10:02:36 | 1 | So, again, given the complexities of the pricing |
| 10:02:38 | 2 | schemes by vendors, I really couldn't give an opinion on that. |
| 10:02:42 | 3 | Q. Right. But you would expect if I had two terabytes of |
| 10:02:44 | 4 | ESI, it would cost more to process that for supervised |
| 10:02:48 | 5 | learning than if I had one terabyte of ESI? |
| 10:02:50 | 6 | A. Well, as I said, if the two terabytes were a couple of |
| 10:02:54 | 7 | long visual video files and the one terabyte was millions of |
| 10:02:58 | 8 | email messages, then it wouldn't necessarily work out that |
| 10:03:02 | 9 | way. One simply can't do a comparison on that level. |
| 10:03:06 | 10 | Q. By the way, your role with Kroll as a consultant, do you |
| 10:03:10 | 11 | have any ownership interest in Kroll? |
| 10:03:12 | 12 | A. No, I do not. |
| 10:03:14 | 13 | Q. It wasn't clear to me last time from your testimony. Have |
| 10:03:16 | 14 | you written one algorithm for Kroll or a number of algorithms |
| 10:03:22 | 15 | for them? |
| 10:03:22 | 16 | A. I have written several algorithms for them. |
| 10:03:26 | 17 | Q. And do you receive any type of royalty if your algorithms |
| 10:03:28 | 18 | are used in terms of when Kroll applies its software? |
| 10:03:32 | 19 | A. No, there's no royalties and I have no royalty or |
| 10:03:36 | 20 | ownership interest in any patents. Those are all Kroll's |
| 10:03:38 | 21 | ownership. |
| 10:03:38 | 22 | Q. Let's go back to Plaintiffs' Exhibit 10. I'd like to go |
| 10:03:42 | 23 | to step 3, which is review the test set. |
| 10:03:46 | 24 | A. Okay. |
| 10:03:48 | 25 | Q. And what's happening in step 3, and please correct me if |

10:03:56   1   I'm wrong, but as I understand step 3, what's going to happen

10:04:00   2   is you have taken this sample from the universe in step 2 to

10:04:06   3   do this test set creation, and then you're going to have

10:04:08   4   someone review it and mark the documents as responsive or

10:04:12   5   nonresponsive; is that correct?

10:04:12   6   A.  That's correct.

10:04:14   7   Q.  And for the indirect method that you described above with

10:04:18   8   these 2,401 documents, you want some senior attorney to do

10:04:24   9   that review, correct?

10:04:24  10   A.  That's what the protocol specifies.

10:04:28  11   Q.  Well, is that your view that it should be a senior

10:04:30  12   attorney that does that review?

10:04:32  13   A.  I do not express an opinion on the type of personnel who

10:04:36  14   would do the review.  I express an opinion that they should be

10:04:42  15   people who can carry out the review with a very high degree of

10:04:44  16   accuracy.

10:04:46  17   Q.  Now, when we look to step 4, called Initialized Training

10:04:56  18   Set on Plaintiffs' Exhibit 10, we now have a training set, and

10:05:00  19   that is different from the test set, correct?

10:05:04  20   A.  That's correct.

10:05:04  21   Q.  All right.  And the training set is this group of

10:05:08  22   documents that is going to be fed into the computer to help

10:05:10  23   train the computer as to what is viewed as responsive or

10:05:14  24   nonresponsive; is that correct?

10:05:16  25   A.  Right.  So in step 4, this is initializing the first

10:05:18   1    version of that training set that the algorithm uses to try to

10:05:24   2    determine how to distinguish responsive and nonresponsive

10:05:28   3    documents.

10:05:28   4    Q.  And then when we turn over to step 5 on the next page of

10:05:32   5    Exhibit 10, where it says, Training review software, do you

10:05:40   6    see that step 5?

10:05:40   7    A.  I see that, yes.

10:05:42   8    Q.  All right.  Now, when we say training review software,

10:05:50   9    basically, you are having the software look at this training

10:05:52   10    set, correct?

10:05:52   11    A.  Correct.

10:05:54   12    Q.  Based on the algorithms that are in whatever system you

10:05:56   13    are using?

10:05:56   14    A.  Right, the algorithms of the training set.

10:06:00   15    Q.  And as a result then is what's going to happen is that the

10:06:04   16    computer is going to make a prediction for the rest of the

10:06:08   17    collection as to what is responsive and nonresponsive?

10:06:12   18    A.  So typically the algorithm will produce a model, a

10:06:18   19    statistical model, and that statistical model then can be run

10:06:22   20    back over the entire collection and will typically produce

10:06:26   21    some numeric score which is intended to be higher if the

10:06:30   22    document is more likely to be responsive and lower if it's

10:06:34   23    less likely to be responsive.

10:06:36   24    Q.  I should have asked you in step 4, the training set, how

10:06:40   25    large is the training set?

| 10:06:42 | 1 | A.  That's something which can vary.  Different vendors will |
| 10:06:46 | 2 | have different best practices for what they consider to be an |
| 10:06:52 | 3 | appropriately sized training set.  So that's something that I |
| 10:06:56 | 4 | left unspecified here because that's really something that |
| 10:07:00 | 5 | should be discussed with the vendor. |
| 10:07:02 | 6 | Q.  All right.  So we have taken this training set, we put the |
| 10:07:08 | 7 | documents in, the algorithm runs, the computer then says, |
| 10:07:16 | 8 | according to what the computer has seen from this training |
| 10:07:18 | 9 | set, the computer is going to rank things in terms of whether |
| 10:07:22 | 10 | it thinks things are responsive or nonresponsive, correct? |
| 10:07:24 | 11 | A.  Yes.  So, typically, when it makes the prediction, it |
| 10:07:28 | 12 | assigns a score to each document, the software would have an |
| 10:07:32 | 13 | option to rank so it can see the highest scoring documents |
| 10:07:36 | 14 | first. |
| 10:07:36 | 15 | Q.  Now, once it has done that, there is a step 6 here on |
| 10:07:40 | 16 | Plaintiffs' Exhibit 10, which says, Find and review additional |
| 10:07:42 | 17 | documents.  And it says, Use the training system as specified |
| 10:07:48 | 18 | by vendors' best practices to find additional unreviewed |
| 10:07:52 | 19 | documents. |
| 10:07:52 | 20 | And by "unreviewed," I assume that means documents |
| 10:07:56 | 21 | that have not yet been reviewed by an attorney, correct? |
| 10:07:58 | 22 | A.  It would be documents that still -- that have not been |
| 10:08:08 | 23 | reviewed for production.  I don't want to say whether it's an |
| 10:08:12 | 24 | attorney or a paralegal or whoever is doing the review for |
| 10:08:14 | 25 | production. |

10:08:14  1  Q.  Hadn't been reviewed by human eyes.  It's only been looked

10:08:18  2  at by the computer so far?

10:08:18  3  A.  Well, the sample, remember that there is a sample that's

10:08:22  4  been drawn, and the sample is being reviewed separately.  And

10:08:28  5  so it is possible that the system will bring up documents that

10:08:32  6  are included in the sample.  And indeed, that's part of the

10:08:36  7  point.  Part of what one is using the sample for is to

10:08:40  8  estimate the effectiveness of the system at finding responsive

10:08:44  9  documents.  To do that, you indeed need to have the system

10:08:48  10  finding some of the sample documents, and those sample

10:08:52  11  documents will have been separately reviewed.

10:08:54  12  Q.  Okay.  Let me step back, because I am trying to understand

10:09:00  13  this process.

10:09:00  14       We have our training set.  We put it into the

10:09:04  15  computer to train the computer.  The computer then codes

10:09:08  16  everything.  Then we are going to find and review additional

10:09:12  17  documents?

10:09:14  18  A.  Right.

10:09:14  19  Q.  And when I look at the third line of step 6 where it says,

10:09:20  20  Review those documents, do you see that line, Dr. Lewis?

10:09:24  21  A.  Yes.

10:09:24  22  Q.  What I want to understand is what are these documents that

10:09:28  23  are being reviewed after I've gone through this initial

10:09:34  24  training set, put it into the computer, had the computer code

10:09:38  25  them, and then I am supposed to find and review additional

10:09:40   1    documents?  What documents am I reviewing?

10:09:44   2    A.  Well, that would depend a bit on vendor's best practices.

10:09:46   3    A common strategy would be to review documents which are at

10:09:52   4    the top of the ranking, to review the documents which are most

10:09:56   5    -- the system currently believes are most likely to be

10:10:00   6    responsive.

10:10:00   7           However, that's not the only possible strategy.  So,

10:10:06   8    for instance, there are systems which use something called

10:10:08   9    active learning, which attempts to show -- as well as

10:10:12  10    top-ranked documents, attempts to show borderline documents

10:10:16  11    that the algorithm believes will be most informative for

10:10:20  12    improving the algorithm.  So there are several possibilities

10:10:22  13    there for which one could get reviewed.

10:10:26  14    Q.  Can you tell me how many documents I am going to have to

10:10:30  15    review in step 6?

10:10:32  16    A.  That's largely a choice of the -- an administrative choice

10:10:38  17    in terms of how many reviewers you have, how long the training

10:10:42  18    function of the system takes.  Again, this is something that

10:10:48  19    one could take into account vendor best practices of.

10:10:52  20           The protocol here doesn't depend on any particular

10:10:54  21    number of documents to be reviewed in step 6 because the

10:10:58  22    protocol terminates when a sufficient level of recall has been

10:11:04  23    achieved.  That might be achieved by doing a small number of

10:11:08  24    very large batches or a large number of very small batches.

10:11:12  25           Again, the intent here is to leave flexibility in the

| | | |
|---|---|---|
| 10:11:16 | 1 | protocol to be able to follow the best practices of any |
| 10:11:20 | 2 | particular vendor of these technologies. |
| 10:11:22 | 3 | Q.  Before we go to step 7, which is estimate the recall, is |
| 10:11:32 | 4 | there any way to know how many documents I am going to have to |
| 10:11:34 | 5 | look at in step 6? |
| 10:11:36 | 6 | A.  You get to choose. |
| 10:11:38 | 7 | Q.  So I could choose, for example, a sample of 665 documents? |
| 10:11:44 | 8 | A.  Well, one -- I don't know if you're confusing the notion |
| 10:11:50 | 9 | of the random sample with the documents that have been found |
| 10:11:54 | 10 | for review.  If, for some reason, you wanted to choose 660 |
| 10:11:58 | 11 | documents in step 6, you could. |
| 10:12:02 | 12 | Q.  Well, I'm sorry, Dr. Lewis, if I'm being obtuse, but what |
| 10:12:06 | 13 | I'm trying to understand, let's go past step 6.  Step 7 is I |
| 10:12:12 | 14 | estimate the recall, and then step 8 is I evaluate the status |
| 10:12:14 | 15 | of the search, and depending on what I evaluate based on my |
| 10:12:18 | 16 | estimate of recall, I determine whether I go back to step 6 |
| 10:12:22 | 17 | and do more of this training; is that correct? |
| 10:12:26 | 18 | A.  Yes.  You're going to look at recall and obviously |
| 10:12:30 | 19 | estimated recall and any other relevant factors here. |
| 10:12:34 | 20 | I should say that while we describe -- discussed |
| 10:12:40 | 21 | returning to step 6, one, depending on, again, best practices, |
| 10:12:44 | 22 | could actually go -- if you had to, you could go all the way |
| 10:12:48 | 23 | back to step 4 and do a seed set, you could go back to step 5 |
| 10:12:52 | 24 | and retrain, or you could go back to step 6 if it's already |
| 10:12:54 | 25 | doing a good job of finding responsive documents and just kind |

| | | |
|---|---|---|
| 10:12:58 | 1 | of review more.  So that probably should have been made clear |
| 10:13:00 | 2 | that there's that additional flexibility here. |
| 10:13:02 | 3 | Q.  For some vendors, is it the case that you take a sample, |
| 10:13:06 | 4 | you review the sample, retrain the system, the computer |
| 10:13:10 | 5 | recodes the documents, and then you pull another sample and |
| 10:13:12 | 6 | continue again? |
| 10:13:12 | 7 | A.  Yes.  There are some vendors who use that strategy where |
| 10:13:18 | 8 | essentially what they do is they choose a random sample, they |
| 10:13:22 | 9 | use it for evaluation.  If the estimated effectiveness is not |
| 10:13:26 | 10 | high enough, they throw that random sample into the training |
| 10:13:30 | 11 | set, retrain, and then draw a new random sample.  So they are |
| 10:13:36 | 12 | always using -- having used up all their old random samples as |
| 10:13:42 | 13 | training data and using only the last random sample for |
| 10:13:44 | 14 | evaluation.  So that, indeed, is a strategy that I am aware at |
| 10:13:48 | 15 | least one vendor follows. |
| 10:13:50 | 16 | Q.  Which vendor is that? |
| 10:13:50 | 17 | A.  That's kCura, I believe.  I'd have to go back and look at |
| 10:13:56 | 18 | the details, but at the high level that I have seen |
| 10:13:58 | 19 | presentations from them, I believe that that's the strategy |
| 10:14:00 | 20 | that they follow.  Now, again, they are always releasing new |
| 10:14:04 | 21 | technological updates, that may have changed, but I have seen |
| 10:14:08 | 22 | some discussion of that. |
| 10:14:08 | 23 | Q.  Does Kroll follow that strategy? |
| 10:14:12 | 24 | A.  I believe that that is not the strategy that they were |
| 10:14:14 | 25 | using.  I believe that that is not the strategy that Kroll is |

10:14:30  1  using.

10:14:30  2  Q.  So that if someone used Kroll rather than kCura, what

10:14:34  3  would they have to do?

10:14:36  4  A.  Again, I don't -- I designed algorithms for Kroll.  I am

10:14:40  5  not aware of all of the details of the operational protocols

10:14:46  6  that they use with their clients, so I would not want to be

10:14:52  7  opining on exactly what they are recommending to their clients

10:14:56  8  right now.

10:14:58  9  Q.  I saw in your materials that you had a reference to the --

10:15:10  10  I think it's the Moore Da La Silva case?

10:15:14  11  A.  Could you point that out?

10:15:14  12  Q.  Sure.  I'm sorry, I gave the case name wrong.  Da Silva

10:15:24  13  Moore.  It's on the fourth page of Plaintiffs' Exhibit 10

10:15:28  14  under note C-1.

10:15:30  15  A.  Yes.

10:15:30  16  Q.  All right.  The Da Silva Moore case, have you looked at

10:15:34  17  the protocol in that case?

10:15:36  18  A.  I have looked at some of the discussion in that case.  I

10:15:42  19  have not looked at it in detail.  The only part that I noted

10:15:46  20  was that they were -- they had a similar problem of drawing a

10:15:52  21  random sample to estimating a proportion in a population, so I

10:15:58  22  thought that would be an informative example for the legal

10:16:00  23  people.

10:16:00  24  Q.  But as you understand how the protocol worked there, it

10:16:04  25  would draw a random sample, train the system, they train the

| | | |
|---|---|---|
| 10:16:08 | 1 | system, the computer would classify all the documents, they |
| 10:16:12 | 2 | would draw a random sample from that, look at the random |
| 10:16:18 | 3 | sample to see how much the computer got right, how much the |
| 10:16:20 | 4 | computer got wrong? |
| 10:16:22 | 5 | A.  I have not looked at the protocol in that case in |
| 10:16:24 | 6 | sufficient detail to give a comment on it.  It's kind of |
| 10:16:26 | 7 | spread out over -- there's a lot of testimony, and I couldn't |
| 10:16:32 | 8 | give you any opinion on that. |
| 10:16:32 | 9 | Q.  Did you speak on the topic of the Da Silva Moore case two |
| 10:16:40 | 10 | weeks ago at the Standard Club? |
| 10:16:40 | 11 | A.  No, I did not. |
| 10:16:42 | 12 | Q.  You did not. |
| 10:16:42 | 13 | Were you on a program at the Standard Club two weeks |
| 10:16:46 | 14 | ago? |
| 10:16:46 | 15 | A.  Yes, I was. |
| 10:16:46 | 16 | Q.  What was the topic of your discussion? |
| 10:16:48 | 17 | A.  Predictive coding. |
| 10:16:50 | 18 | Q.  Now, at the end of -- when you estimate recall at step 7, |
| 10:17:02 | 19 | and I'm not sure how you get out of step 6, but when you |
| 10:17:04 | 20 | estimate recall at step 7, you're going to try to figure out |
| 10:17:08 | 21 | what percentage you have for recall at that point; is that |
| 10:17:10 | 22 | correct?  That's a statistical calculation? |
| 10:17:12 | 23 | A.  Right.  So you are attempting -- you are estimating a |
| 10:17:16 | 24 | confidence interval on recall. |
| 10:17:18 | 25 | Q.  And it could be that the estimate of recall you obtain |

10:17:24 1 after you run through this whole process the first time is you

10:17:28 2 can find out that your recall estimate is 13 percent, plus or

10:17:32 3 minus 5 percent, correct?

10:17:34 4 A. That could be an outcome, yes.

10:17:38 5 Q. And if it were 13 percent after that first round of going

10:17:42 6 through, would you expect to go back through the round again?

10:17:44 7 A. The recall target is something that would have to be

10:17:50 8 determined on a legal basis. There is no right answer from

10:17:56 9 informational retrieval about what a sufficient level of

10:18:00 10 recall is.

10:18:00 11 Q. So 13 percent might be okay?

10:18:02 12 A. I just don't have an opinion on that.

10:18:04 13 Q. Now, in the portion of the Da Silva case that you looked

10:18:14 14 at, did you see that there was a suggestion that the parties

10:18:16 15 had to go through the training and the rounds of this

10:18:26 16 predictive coding they were calling it there processed seven

10:18:28 17 times and then come back to the judges to see if they needed

10:18:30 18 to go through more rounds?

10:18:32 19 A. I didn't look at that part of the protocol. I really just

10:18:34 20 saw this bit about sampling, and that was a good example

10:18:38 21 there.

10:18:38 22 Q. But sitting here today, we could go through this entire

10:18:46 23 protocol and you can't tell me what would be the recall that

10:18:50 24 we would result with after running this process; isn't that

10:18:50 25 correct?

| | | |
|---|---|---|
| 10:18:56 | 1 | A. If you set a recall target and you run the process, you |
| 10:19:02 | 2 | iterate enough times, eventually, you will reach that recall |
| 10:19:08 | 3 | simply because you will have looked at enough documents that |
| 10:19:12 | 4 | it would be inevitable you reach that recall. |
| 10:19:14 | 5 | Q. But you can't tell me how long that iterative process is |
| 10:19:18 | 6 | going to take; is that correct? |
| 10:19:20 | 7 | A. That's correct. |
| 10:19:20 | 8 | Q. And you can't tell me how many times I'm going to have to |
| 10:19:22 | 9 | run through this process; isn't that correct? |
| 10:19:24 | 10 | A. That's correct. |
| 10:19:24 | 11 | Q. I mean, eventually, I would have, God forbid, the senior |
| 10:19:28 | 12 | attorney have to review every document in the entire |
| 10:19:30 | 13 | production or entire universe, correct? |
| 10:19:32 | 14 | A. I don't believe -- I have not expressed an opinion about |
| 10:19:40 | 15 | what sort of personnel would be reviewing these documents. |
| 10:19:42 | 16 | Q. I'm sorry. That was in the protocol. |
| 10:19:44 | 17 | But, I mean, eventually, you say you are going to |
| 10:19:46 | 18 | look at enough documents. You don't know if we are going to |
| 10:19:48 | 19 | be looking at any particular percentage of the documents or |
| 10:19:52 | 20 | how high a percentage of the documents, correct? |
| 10:19:52 | 21 | A. No, my opinion is that using supervised learning and |
| 10:19:58 | 22 | statistical rank retrieval, you will have to look at fewer |
| 10:20:00 | 23 | documents than if you used Boolean queries. |
| 10:20:04 | 24 | Q. And you can't tell me what it's going to cost? |
| 10:20:06 | 25 | A. No, I can't. |

10:20:06 1    Q.  All right.  I want to try to move a little bit to sort of

10:20:14 2    inside how this system works, and you talked about this some

10:20:18 3    on direct.  But this computer algorithm that we have talked

10:20:22 4    about, it's still relying, in part, on words, correct?

10:20:26 5    A.  Supervised learning and statistical rank retrieval

10:20:32 6    algorithms can make use of words in text, phrases in text,

10:20:36 7    metadata in the text.  All of those things can go into the

10:20:40 8    algorithm.

10:20:40 9    Q.  And the algorithm is going to distinguish responsive

10:20:44 10   documents from nonresponsive documents by giving various

10:20:48 11   weights to different terms, metadata, phrases, correct?

10:20:52 12   A.  Yes.  Typically, all the algorithms that I am aware of

10:20:56 13   that are being used for e-discovery produce some sort of

10:21:02 14   weighting of the different pieces of evidence.

10:21:06 15   Q.  And as part of this process, is there some type of word

10:21:08 16   index that the computer creates up front so it knows what

10:21:14 17   terms it's going to be looking at or considering?

10:21:16 18   A.  That would be a common way to implement such systems is to

10:21:22 19   produce an index of both the text and the metadata.  For

10:21:26 20   instance, you know, which documents contain which words and

10:21:30 21   metadata, how many times they occur, sometimes positional

10:21:36 22   information.

10:21:36 23   Q.  By the way, how long does this computer program take to

10:21:38 24   run?

10:21:38 25   A.  That would depend on the particular implementation.

10:21:40　1　Q.　If you had, say, a million documents in the program, how

10:21:44　2　long would it take for the algorithm to work through them all

10:21:48　3　the first time?

10:21:48　4　A.　How long -- you are asking how long the indexing would

10:21:52　5　take?

10:21:52　6　Q.　Well, let's start with the indexing, yes.

10:21:54　7　A.　I mean, again, it's something which is going to depend on

10:21:58　8　the particular vendor software.　Indexing a million documents

10:22:06　9　is something which is not a very large task.　I certainly, you

10:22:12　10　know, am aware of other -- of search software that can do that

10:22:18　11　in less than a minute.

10:22:20　12　Q.　Less than a minute?

10:22:22　13　A.　Yeah.

10:22:22　14　　　　　THE COURT:　A million?　A million documents in one

10:22:24　15　minute?

10:22:24　16　　　　　THE WITNESS:　I mean, you know, there's distributive

10:22:26　17　search software that runs across multiple computers.　The

10:22:30　18　indexing is not a substantial difficult culling.

10:22:36　19　BY MR. McKEOWN:

10:22:36　20　Q.　But at the end of this indexing, what the computer has

10:22:38　21　generated are all these different things it's going to look at

10:22:44　22　in a way; is that correct?

10:22:44　23　A.　Yeah.

10:22:46　24　Q.　There will be a bunch of words on this list, right?

10:22:48　25　A.　It goes through the documents, and it extracts information

10:22:50    1   on the piece of evidence that's going to be used.  It could be

10:22:54    2   words, phrases, metadata, and so on.

10:22:56    3   Q.  And isn't it also true that the more common or frequent

10:23:08    4   the term, the more difficult it's going to be for the computer

10:23:10    5   or the algorithm to sort between responsive and nonresponsive?

10:23:14    6   A.  I would phrase that slightly differently.  Very frequent

10:23:20    7   terms, you think of, say, common English words, the, a, and

10:23:26    8   whatnot, tend not to be very discriminative between

10:23:30    9   responsiveness and nonresponsiveness, so they don't tend to

10:23:34   10   get very much weight when the computer is discriminating

10:23:38   11   between the two.

10:23:40   12   Q.  Okay.  Let me see if I understand that.

10:23:42   13          When you say "discriminating between the two," you

10:23:46   14   mean when the computer is classifying it as responsive or

10:23:48   15   nonresponsive?

10:23:50   16   A.  Right.  So these very common terms don't tend to be as

10:23:54   17   helpful in classifying something as responsive versus

10:23:58   18   nonresponsive as a more moderate frequency piece of evidence.

10:24:06   19   Q.  Now, we had talked at the last session about the TREC 2009

10:24:22   20   conference, and I just want to go back to that briefly.

10:24:26   21          Did you, by any chance, go back and check the TREC

10:24:30   22   2009 conference report since we last were here?

10:24:32   23   A.  I don't think I've actually looked at TREC 2009 since

10:24:36   24   then.

10:24:36   25   Q.  I want to show you what was previously marked as

10:24:54    1   Defendants' Exhibit 7.

10:24:56    2           THE COURT:  Did we admit that?

10:25:02    3           Thank you.  Did you want to move to admit it?

10:25:12    4           MR. McKEOWN:  Not at this stage, your Honor.  I want

10:25:14    5   to talk about it just a little first.

10:25:18    6           THE COURT:  Okay.

10:25:18    7   BY MR. McKEOWN:

10:25:20    8   Q.  And you recall this document, Dr. Lewis?

10:25:22    9   A.  Yes, I do.

10:25:24   10   Q.  And we talked about this in the context of the TREC 2009

10:25:30   11   legal track had used the Enron emails for their source data,

10:25:34   12   correct?

10:25:34   13   A.  I know that they used Enron for some of it.  However, I

10:25:42   14   believe they also used the INT feedback fee portion of the

10:25:48   15   tobacco documents for another part of it.

10:25:50   16   Q.  Let me just direct your attention -- by the way, if I ever

10:25:54   17   make a contribution to information retrieval, I'd ask that

10:26:00   18   they put page numbers on their publications.  But it's the

10:26:02   19   fourth page --

10:26:02   20   A.  Please talk to Ness (phonetic) about that.

10:26:04   21   Q.  It's under Section 2.2.1.

10:26:08   22   A.  Okay.

10:26:14   23   Q.  If you go down to the fifth line, it says, We turned in

10:26:22   24   2009 to a collection of emails that had been produced by Enron

10:26:26   25   in response to requests from the Federal Energy Regulatory

10:26:28  1 | Commission, correct?

10:26:30  2 | A.  I see that, yes.

10:26:30  3 | Q.  Okay.  So it appears that, at least for this particular

10:26:34  4 | year --

10:26:34  5 | A.  Right.

10:26:34  6 | Q.  -- they were using the Enron emails.

10:26:36  7 | A.  Right.  They used Enron for one of the tasks, and they

10:26:38  8 | used tobacco for another one of the tasks.

10:26:42  9 | Q.  And we talked the last time about one of the topics that

10:26:46  10 | they looked at here when they ran their request was a request

10:26:48  11 | for an effort to find things about fantasy football and

10:26:54  12 | gambling on football.  Do you recall that?

10:26:56  13 | A.  I remember you asking about that.

10:26:58  14 | Q.  Well, if you turn to what is the sixth page, at the top of

10:27:08  15 | the page is a list for topic 204.

10:27:12  16 | A.  Okay.

10:27:14  17 | Q.  Have you found that?

10:27:14  18 | A.  Yes, I see topic 204.

10:27:16  19 | Q.  Okay.  And then it's right above Section 2.2.3, there's

10:27:22  20 | topic 207.

10:27:24  21 | A.  Yes, I see it.

10:27:24  22 | Q.  And that says what they are looking for is the document

10:27:28  23 | request are, All documents or communications that describe,

10:27:30  24 | discuss, refer to, report on, or relate to fantasy football,

10:27:34  25 | gambling on football, and related activities, including but

| | | |
|---|---|---|
| 10:27:38 | 1 | not limited to football teams, football players, football |
| 10:27:42 | 2 | games, football statistics, and football performance, correct? |
| 10:27:46 | 3 | A.  Yes. |
| 10:27:48 | 4 | Q.  For this particular request, that's what the teams were |
| 10:27:50 | 5 | supposed to look at, correct? |
| 10:27:50 | 6 | A.  Yes. |
| 10:27:52 | 7 | Q.  And we talked the last time about only five of the -- let |
| 10:28:00 | 8 | me get the term -- was it runs, 24 runs? |
| 10:28:04 | 9 | A.  Could you point me to the area -- |
| 10:28:08 | 10 | Q.  Sure.  If we go to -- I had the benefit last night of |
| 10:28:12 | 11 | writing numbers in the corners of my pages, but it is the 16th |
| 10:28:14 | 12 | page.  It's right above Section 2.3.6. |
| 10:28:24 | 13 | A.  Okay.  I see the page that has Section 2.3.6. |
| 10:28:28 | 14 | Q.  And if you look at the paragraph that's about 5 inches |
| 10:28:36 | 15 | from the bottom of that page that starts, As can be seen from |
| 10:28:40 | 16 | the table? |
| 10:28:40 | 17 | A.  Um-hmm. |
| 10:28:44 | 18 | Q.  Do you see that paragraph, Dr. Lewis? |
| 10:28:46 | 19 | A.  I see that, As can be seen from the table. |
| 10:28:48 | 20 | Q.  All right.  And drop down to the fourth line in about an |
| 10:28:50 | 21 | inch, it says, In terms of recall, of the 24 submitted runs, |
| 10:28:56 | 22 | five distributed across four topics attained a recall score of |
| 10:29:00 | 23 | 0.7 or greater. |
| 10:29:02 | 24 | Do you see that line? |
| 10:29:04 | 25 | A.  Yes. |

| | | |
|---|---|---|
| 10:29:04 | 1 | Q. And in that four topics, two of those topics were topic |
| 10:29:10 | 2 | 207; isn't that right? |
| 10:29:12 | 3 | A. Yes, 207 UW and 207 CB are the runs. |
| 10:29:16 | 4 | Q. And so of the five that reach recall of 0.7 or greater, |
| 10:29:22 | 5 | two of them were our football requests? |
| 10:29:24 | 6 | A. That's correct. |
| 10:29:24 | 7 | Q. Correct? |
| 10:29:24 | 8 | A. Yes. |
| 10:29:24 | 9 | Q. And, in fact, if we turn to the very next page, we'd have |
| 10:29:30 | 10 | table 6, which fills up the first top half of the page there. |
| 10:29:36 | 11 | Do you see that? |
| 10:29:36 | 12 | A. Yes. |
| 10:29:36 | 13 | Q. And looking at the columns, the first column says, Topic, |
| 10:29:42 | 14 | the second column is Run, the third column is actually three |
| 10:29:46 | 15 | columns together under the Recall header. |
| 10:29:48 | 16 | Do you see that? |
| 10:29:48 | 17 | A. I do. |
| 10:29:48 | 18 | Q. And where it says, EST, period, that's the estimate of |
| 10:29:52 | 19 | recall; is that right? |
| 10:29:54 | 20 | A. That's my understanding. |
| 10:29:54 | 21 | Q. And if we go down to topic 207, we see UW and CB were the |
| 10:30:00 | 22 | two teams that had recall over 0.7, correct? |
| 10:30:04 | 23 | A. Yes, on topic 207, there were those two teams. |
| 10:30:08 | 24 | Q. And if you look at topic 206, which is just above it on |
| 10:30:12 | 25 | that table? |

10:30:14    1   A.   Yes.

10:30:14    2   Q.   Now, do I understand this correct that the first team, CB

10:30:18    3   high, had 0.076 recall?

10:30:22    4   A.   That's my understanding.

10:30:24    5   Q.   So that would be 7.6 percent recall, correct?

10:30:28    6   A.   That's correct.

10:30:28    7   Q.   So that CB high on topic 206, when they ran their process,

10:30:34    8   they only got 7.6 percent of the responsive documents,

10:30:38    9   correct?

10:30:40   10   A.   That's my understanding.

10:30:40   11   Q.   And then there's something, a team LO, and they've got 4.2

10:30:46   12   percent; is that correct?

10:30:48   13   A.   That's correct.

10:30:48   14   Q.   And CB mid had 1.1 percent recall; is that correct?

10:30:54   15   A.   Yeah.  My understanding is that CB is a single team and

10:30:56   16   that high, mid, and LO would have been three runs that they

10:31:00   17   submitted.

10:31:00   18   Q.   Okay.  So CB LO is 0.9 percent?

10:31:04   19   A.   Right.

10:31:04   20   Q.   So their low run, was that low-by-low score or low-by-low

10:31:10   21   effort?

10:31:12   22   A.   I don't know how they're coding their runs.

10:31:18   23   Q.   But, clearly, the recall return on top 206 is not so

10:31:22   24   great.  Would you agree with that?

10:31:24   25   A.   I would agree that it's substantially lower than the

10:31:28    1    recall on topic 207.

10:31:28    2    Q.   And if we go back to the sixth page where we were just a

10:31:32    3    few minutes ago which has the list of topics?

10:31:38    4    A.   Okay.

10:31:38    5    Q.   This is right above Section 2.2.3.  When we look at topic

10:31:46    6    206 -- and, again, we are looking at the Enron emails, right?

10:31:52    7    A.   That's right.

10:31:52    8    Q.   So topic 206 was all documents or communications that

10:31:56    9    describe, discuss, refer to, report on, or relate to any

10:32:02   10    discussions, communications, or contacts with financial

10:32:04   11    analysts or with firms that employ them regarding the

10:32:06   12    company's financial condition, analyst's coverage of the

10:32:10   13    company and/or its financial condition, analyst's rating of

10:32:14   14    the company's stock, or the impact of an analyst's coverage on

10:32:16   15    the company -- on the business relationship between the

10:32:20   16    company and the firm that employs the analyst.

10:32:22   17         Do you see that?

10:32:22   18    A.   I do.

10:32:22   19    Q.   So on football, they did great.  On requests about

10:32:28   20    financial analysts, the performance was pretty poor.  Would

10:32:32   21    you agree?

10:32:32   22    A.   I would agree that the performance in terms of recall was

10:32:34   23    substantially lower on topic 206.

10:32:36   24    Q.   It was about 10 percent of what they got on football,

10:32:40   25    correct?

10:32:40   1   A.   Yes, 10 percent of the recall on the item.

10:32:48   2   Q.   Now, when TREC 2009 ran their legal track, these teams ran

10:32:52   3   on -- you know, they would run a search for a single request;

10:32:56   4   is that correct?

10:32:56   5   A.   Could you describe a little more precisely what you mean?

10:33:00   6   Q.   Each team could take one, two, or three requests, but they

10:33:06   7   were running a process aimed for that one focused request,

10:33:16   8   correct?

10:33:18   9   A.   When you say "request," you mean the topics?

10:33:18   10   Q.   The topic. I'm sorry. I'm using a legal term in requests

10:33:22   11   for production.

10:33:24   12           But they were running -- when they ran the topic,

10:33:26   13   they were searching just for football?

10:33:28   14   A.   I don't know the details of what the particular teams did.

10:33:32   15   Q.   Would you agree that if you had to search for a number of

10:33:34   16   topics across a universe of documents, it would be more

10:33:38   17   difficult for the supervised learning process to cover all

10:33:42   18   those topics?

10:33:42   19   A.   That would depend on the details of the topics.

10:33:46   20   Q.   If the topics were diverse, it would take longer and more

10:33:52   21   effort to locate the documents with each topic; is that true?

10:33:54   22   A.   That's something one would really have to get into the

10:33:58   23   details. It would depend on the details of the requests, it

10:34:06   24   would depend on what one was seeking in terms of the

10:34:10   25   effectiveness, whether one was seeking a high level of

10:34:16  1  effectiveness, for instance, on the union of the responsive

10:34:20  2  documents from the set of them or if there were individual

10:34:22  3  effectiveness targets for each of the requests.

10:34:26  4        So one would really have to understand the details of

10:34:30  5  the need in that operational situation.

10:34:32  6  Q. Would you agree that you would have a richer training set

10:34:38  7  if you drew a sample for your training set that was entirely

10:34:42  8  devoted to football than if you drew a training set of the

10:34:46  9  same number of documents but it covered 29 different

10:34:50  10  categories of which football was only one?

10:34:52  11  A. Could you repeat the question?

10:34:54  12  Q. Sure. Let's say we have a training set of 500 documents.

10:35:00  13  If I am only searching for one topic, football, all 500

10:35:04  14  documents are going to talk about football, right?

10:35:06  15  A. Well, typically in training sets, you want examples of

10:35:10  16  both responsive and nonresponsive documents.

10:35:12  17  Q. Fair point.

10:35:14  18        But the distinction in all those documents is going

10:35:18  19  to be football or no football if I am only looking for

10:35:20  20  football requests, correct?

10:35:22  21  A. Again, I'm not sure what you are saying there. A training

10:35:28  22  set -- what's important about a training set is you are trying

10:35:32  23  to detect documents about football is that you have some

10:35:36  24  documents about football and some that aren't about football

10:35:38  25  and that documents have been reviewed to label them in an

10:35:42  1  unbiased fashion with those labels.

10:35:44  2  Q.  But those 500 documents, that's the only thing they'd be

10:35:48  3  labeled for, correct?

10:35:48  4  A.  If that's the only goal that you have is to detect topics

10:35:52  5  about football.

10:35:54  6  Q.  If you had 30 different topics that I was looking at, I

10:35:56  7  would have to label those documents for all those topics,

10:35:58  8  correct?

10:35:58  9  A.  Again, that would depend on what the goal of the system

10:36:06  10  was.  If you needed separate predictions for each of those 30

10:36:12  11  topics, you would need training data labeled for each of those

10:36:16  12  topics.

10:36:18  13        If you only needed a prediction for whether a

10:36:20  14  document was in the union of those 30 topics, then you might

10:36:26  15  have labeling for the individual topics or you might just have

10:36:30  16  a label for is it in the union or not.

10:36:32  17  Q.  And if I only wanted to do responsiveness, then I'd

10:36:36  18  probably have fewer documents about football in my training

10:36:40  19  set if I had a lot of other topics as well, correct?

10:36:42  20  A.  This would depend what topic responsiveness was with

10:36:46  21  respect to.

10:36:48  22  Q.  You were also asked about, with respect to the training

10:36:52  23  set, whether you could create a document for this training set

10:36:58  24  or seed set with the words "they are with us."  Do you

10:37:02  25  remember that testimony, "They are with us"?

10:37:08   1   A.   I believe there was a question about that.

10:37:10   2   Q.   And I think you said that each of those four words was a

10:37:14   3   high frequency word, correct?

10:37:14   4   A.   That's correct.

10:37:16   5   Q.   And, therefore, when we talked about indexing before,

10:37:20   6   those are the types of words that the system is not likely to

10:37:24   7   put much weight to; is that correct?

10:37:26   8   A.   Right.   Those are words that are difficult to -- they tend

10:37:30   9   to have relatively little value individually in distinguishing

10:37:34   10  documents as responsive or not responsive.

10:37:36   11  Q.   And, therefore, it could pick up all sorts of things.   If

10:37:42   12  you tried to run them to try to find other documents and said,

10:37:44   13  They are with us, the computer is going to have difficulty

10:37:46   14  sorting those out; isn't that right?

10:37:50   15  A.   Yeah.   I mean, there are cases where, you know, if you

10:37:54   16  knew that systematically that exact string was used, one could

10:37:58   17  turn that exact string into a phrase.   But, in general, to get

10:38:02   18  -- if responsive documents had only high frequency words, it's

10:38:06   19  going to be difficult to distinguish them from nonresponsive

10:38:10   20  documents using only word-based evidence alone.

10:38:12   21  Q.   In fact, if you knew you were looking for that exact

10:38:16   22  phrase, you could use a Boolean search term, couldn't you?

10:38:22   23  A.   If you were using that exact phrase, yes, you could use

10:38:24   24  like proximity operators.

10:38:26   25  Q.   Because just putting in documents, even a made-up document

10:38:30   1   that says, They are with us, it could be things like, They

10:38:34   2   will be meeting us for dinner, or, They will be at breakfast

10:38:40   3   tomorrow, or it could be any sorts of things that might be

10:38:42   4   picked up, correct?

10:38:42   5   A.   But what you do when you are creating a made-up document

10:38:46   6   of that sort is you attempt to also take the metadata that

10:38:50   7   would be characteristic of the document as well because the

10:38:52   8   individual words, as you pointed out, finding the documents

10:38:54   9   purely based on a word search where you don't know what, say,

10:39:00   10   a coded phrase like that would be, you don't want to find just

10:39:04   11   the ones with the un-uncoded phrase because you don't know the

10:39:10   12   un-uncoded phrase.  What you hope to do is create artificial

10:39:14   13   documents that had characteristic metadata as well.

10:39:16   14   Q.   And so you're going to make up a document with both

10:39:20   15   phrases and metadata?

10:39:20   16   A.   Well, the idea one -- this is a possible way -- possible

10:39:26   17   technique could be used in this protocol is to create

10:39:30   18   simulated documents with both text and metadata.

10:39:32   19   Q.   But it's still going to have high frequency words,

10:39:38   20   correct?

10:39:38   21   A.   The words will be high frequency.  Obviously, the metadata

10:39:40   22   will have other characteristics.

10:39:42   23   Q.   And the metadata is going to be a complete guess?

10:39:44   24   A.   Well, it would obviously depend on the insight of the

10:39:48   25   person creating the artificial training document into the --

| | | |
|---|---|---|
| 10:39:52 | 1 | you know, what metadata would be likely to occur in a |
| 10:39:56 | 2 | responsive document. |
| 10:39:56 | 3 | Q.  Let's talk a little bit again about the algorithm.  Now, |
| 10:40:02 | 4 | you wrote some algorithms for Kroll. |
| 10:40:04 | 5 | Do different algorithms act differently when |
| 10:40:08 | 6 | documents are fed in as a training set? |
| 10:40:12 | 7 | A.  Yes, there are a number of different kinds of supervised |
| 10:40:14 | 8 | learning algorithms which will have somewhat different |
| 10:40:18 | 9 | responses in its training data. |
| 10:40:22 | 10 | Q.  When you say somewhat different responses, you mean that |
| 10:40:24 | 11 | different algorithms, if you feed the same documents into |
| 10:40:28 | 12 | them, could code documents differently, correct? |
| 10:40:32 | 13 | A.  Right.  So different algorithms would produce different |
| 10:40:36 | 14 | predictive models, and, you know, a model being what you apply |
| 10:40:40 | 15 | to the document to get a predictive score of relevance.  And |
| 10:40:44 | 16 | different of those models would score documents differently, |
| 10:40:48 | 17 | and so there could be some difference in the predictions of |
| 10:40:50 | 18 | which are responsive and which are nonresponsive. |
| 10:40:54 | 19 | Q.  And you didn't write the algorithm for any of the other |
| 10:40:56 | 20 | companies, correct? |
| 10:40:58 | 21 | A.  No, I have only done algorithms on Kroll. |
| 10:41:00 | 22 | Q.  So that if we took the exact universe of documents, |
| 10:41:06 | 23 | whatever it is we want to put through, and the exact same |
| 10:41:08 | 24 | training set and fed it into the Kroll system with your |
| 10:41:12 | 25 | algorithm and fed it into a different system with a different |

| | | |
|---|---|---|
| 10:41:16 | 1 | algorithm, there is going to be different predictions by those |
| 10:41:20 | 2 | two algorithms; isn't that correct? |
| 10:41:22 | 3 | A.  Yes, one would expect that some of the predictions would |
| 10:41:24 | 4 | be there. |
| 10:41:24 | 5 | Q.  And that's because the algorithms are different? |
| 10:41:28 | 6 | A.  Yes. |
| 10:41:28 | 7 | Q.  You testified at our last session about a study called |
| 10:41:38 | 8 | Blair and Maron.  Do you recall that? |
| 10:41:42 | 9 | A.  Yes. |
| 10:41:42 | 10 | Q.  And, in fact, I think you listed that study as a reference |
| 10:41:46 | 11 | in the report you prepared in this case, correct? |
| 10:41:48 | 12 | A.  Yes. |
| 10:41:48 | 13 | Q.  And I think you said that, in your view, that study had to |
| 10:41:56 | 14 | do with the attorneys thinking they had found 75 percent of |
| 10:42:00 | 15 | the responsive documents and they had actually found a lot |
| 10:42:04 | 16 | lower percentage, correct? |
| 10:42:04 | 17 | A.  Yes. |
| 10:42:06 | 18 | Q.  Now, in that study, wasn't it the case that what happened |
| 10:42:08 | 19 | was there were two lawyers and two paralegals, and the lawyers |
| 10:42:14 | 20 | were asked to tell the paralegal what type of request to |
| 10:42:20 | 21 | create, and the paralegals then queried the system to try to |
| 10:42:24 | 22 | draw back documents? |
| 10:42:24 | 23 | A.  I'd have to look at the original study to verify that. |
| 10:42:58 | 24 | MR. McKEOWN:  Your Honor, may I approach the witness? |
| 10:43:00 | 25 | THE COURT:  Yes. |

10:43:02    1          Did you show Dr. Lewis this the last time --

10:43:06    2          MR. McKEOWN:  No, I did not, your Honor.  This is a

10:43:08    3  new exhibit.  This would be Defendants' Exhibit 9.

10:43:10    4          THE COURT:  All right.

10:43:12    5          THE WITNESS:  Thank you.

10:43:16    6          MR. MOGIN:  I'm sorry.  This is 9?

10:43:18    7          MR. McKEOWN:  I believe so.

10:43:28    8  BY MR. McKEOWN:

10:43:32    9  Q.  Dr. Lewis, you have a copy of Defendants' Exhibit 9?

10:43:34   10  A.  Yes, I do.

10:43:36   11  Q.  Is this the Blair and Maron study that you were

10:43:40   12  referencing before?

10:43:42   13  A.  Yes, it is.

10:43:44   14  Q.  If you turn to page 291 of Defendants' Exhibit 9, at the

10:44:02   15  top of the page where it says, The conduct of the test, do you

10:44:06   16  see that?

10:44:06   17  A.  Yes, I do.

10:44:06   18  Q.  And in the third line, it says, Two lawyers, the principal

10:44:10   19  defense attorneys in the suit, participated in the experiment?

10:44:14   20  A.  Yes, I see that.

10:44:14   21  Q.  Does that refresh your recollection as to how that was

10:44:16   22  done?

10:44:16   23  A.  I see that that says that the two lawyers participated in

10:44:24   24  an experiment.

10:44:26   25  Q.  And then you will see, if you look down, that they

10:44:30  1  generated a total of 51 different information requests which

10:44:34  2  were translated into formal queries by either of the two

10:44:38  3  paralegals, both of whom were familiar with the case?

10:44:40  4  A.  I see that.

10:44:40  5  Q.  So it appears that the -- you had two lawyers, they

10:44:46  6  created requests, and then the paralegals created queries to

10:44:50  7  search the database, correct?

10:44:52  8  A.  Yes.

10:44:52  9  Q.  And what happened then was that the lawyers were told to

10:44:58  10 continue this process until they thought they had 75 percent

10:45:02  11 of the documents; is that correct?

10:45:04  12 A.  Until they thought that they had 75 percent of the

10:45:08  13 responsive documents.

10:45:08  14 Q.  I'm sorry, yes, until they thought they had 75 percent of

10:45:12  15 the responsive documents.

10:45:14  16       By the way, if you look at the middle of that same

10:45:16  17 column -- let me do it by lines.

10:45:26  18       About 13 lines down from the top on the right-hand

10:45:30  19 column starting at the end of the line, it says, The lawyer

10:45:34  20 then evaluated the documents ranking them according to whether

10:45:38  21 they were vital, satisfactory, marginally relevant, or

10:45:40  22 irrelevant to the original request; is that right?

10:45:42  23 A.  Yes, I see that.

10:45:44  24 Q.  So at least in this famous study, they listed things and

10:45:54  25 at least categorized some of them as marginally relevant?

10:45:58  1  A.  They did.

10:45:58  2  Q.  So in this case, the end of the study came when the

10:46:00  3  lawyers --

10:46:00  4       MR. MOGIN:  Your Honor, may the witness be asked to

10:46:04  5  read from the rest of the document where the use of

10:46:06  6  "marginally relevant" is explained three sentences down?

10:46:12  7       THE COURT:  I'm sorry.  I was actually looking down,

10:46:14  8  so I missed that.  Is that part of the sentence?

10:46:16  9       MR. MOGIN:  He read from part of the paragraph.  I

10:46:18  10 was wondering if the witness could be asked to read the rest

10:46:22  11 of the paragraph.

10:46:22  12      THE COURT:  Do you mind, Mr. McKeown, rather than

10:46:24  13 doing it on redirect, can you just have him do it now, please?

10:46:30  14      MR. McKEOWN:  That's fine.

10:46:30  15      THE COURT:  Thank you.

10:46:34  16      MR. McKEOWN:  Is there a particular line?

10:46:36  17      THE COURT:  What page?  Tell me what page,

10:46:40  18 Mr. McKeown.  I'm sorry.

10:46:42  19      MR. McKEOWN:  We are on page 291 of the document,

10:46:44  20 your Honor.

10:46:44  21      THE COURT:  And you were on -- halfway under conduct

10:46:48  22 of the test.

10:46:48  23      MR. McKEOWN:  We were 13 lines down.

10:46:50  24      THE COURT:  Okay.  Fine.  So ask Dr. Lewis again the

10:46:54  25 question incorporating Mr. Mogin's comments, if you will.

10:47:04   1   BY MR. McKEOWN:

10:47:06   2   Q. If you look down, we were just on that 13 lines down where

10:47:10   3   it said, The lawyer then evaluated the documents ranking them

10:47:14   4   according to whether they were vital, satisfactory, marginally

10:47:16   5   relevant, or irrelevant to the original request, correct?

10:47:20   6   A. Yes.

10:47:20   7   Q. And then if you skip two sentences, so we are down about

10:47:28   8   six lines, the sentence, the line starts with the word

10:47:32   9   recorded at the end of the last sentence.

10:47:34   10   A. Okay.

10:47:34   11   Q. I want to point to the next one: The information

10:47:38   12   requested in query formulation procedures were considered

10:47:40   13   complete only when the lawyers stated in writing that he or

10:47:44   14   she was satisfied with the search results for that particular

10:47:48   15   query, i.e., in his or her judgment, more than 75 percent of

10:47:52   16   the vital satisfactory and marginally relevant documents had

10:47:56   17   been retrieved; is that right?

10:47:58   18   A. I do see that.

10:47:58   19   Q. So what they did was they ran these queries and the

10:48:04   20   lawyers made that judgment. However, they did no testing of

10:48:08   21   their null set; isn't that correct?

10:48:10   22   A. Well, at this point when they are -- yeah, I mean, they

10:48:16   23   are looking at a series of documents, and they are not --

10:48:20   24   there is no description here of any testing of the null set

10:48:22   25   that went on at that point. I don't know what the paralegals

10:48:24  1  were doing.

10:48:26  2  Q.  They had two lawyers, principal defense lawyers.  Lawyers

10:48:32  3  were asked when you think you have gotten 75 percent of the

10:48:36  4  responsive documents, they formed a judgment, their judgment

10:48:40  5  may have been wrong, but they didn't have anything -- there is

10:48:42  6  nothing in this report that suggests that they ran a null set

10:48:44  7  test, right?

10:48:46  8  A.  There is no description of a null set test, of a test of

10:48:48  9  the end review documents at this point in the process.

10:48:52  10  Obviously later in the paper, they talk about doing the

10:48:54  11  sampling.

10:48:56  12  Q.  To show that the lawyers were wrong about their guess?

10:49:00  13  A.  Yes, yes.

10:49:00  14  Q.  Which would be why you would test the null set to see if

10:49:08  15  you were wrong, correct?

10:49:08  16  A.  Yes.

10:49:08  17  Q.  You also testified, apart from the Blair and Maron study,

10:49:12  18  that you thought that the reviewers that would be looking at a

10:49:16  19  null set sample should not know that it's a null set sample;

10:49:16  20  is that right?

10:49:20  21  A.  That's correct.

10:49:22  22  Q.  And you think that bias would influence their judgment of

10:49:24  23  a document as responsive or nonresponsive?

10:49:28  24  A.  That's certainly possible.

10:49:30  25  Q.  Now, you're not an attorney, correct?

10:49:32  1  A.  That's correct.

10:49:32  2  Q.  Now, is it your belief that the fact that an attorney owes

10:49:34  3  some duty of loyalty to a client means that they are going to

10:49:38  4  grade things differently if they know that it's from the null

10:49:40  5  set and not a null set?

10:49:42  6  A.  I am not expressing an opinion on that.

10:49:44  7  Q.  Okay.  Let's say, for example, that Temple-Inland had

10:49:50  8  Ms. Miller review all the documents in the null set but she

10:49:54  9  knew they were from the null set.  It's your view that because

10:49:58  10  she knew they were from the null set that her judgment on

10:50:00  11  those documents is not to be trusted; is that correct?

10:50:04  12  A.  I'm not making any statement about Ms. Miller right now.

10:50:08  13  I'm simply making a statement about the difficulty of human

10:50:10  14  beings making subjective judgments about the responsiveness of

10:50:16  15  the document making an objective -- making an unbiased

10:50:22  16  judgment of that when they know what a computer system did or

10:50:26  17  didn't do to that document.

10:50:26  18  Q.  And in your view, that's not to be trusted?

10:50:30  19  A.  That's correct.

10:50:32  20       THE COURT:  I have a question then.  Okay.  Well,

10:50:42  21  then who would be -- how would the person doing the review, if

10:50:46  22  it's not somebody knowledgeable in the case or knowledgeable

10:50:52  23  of what was either supposed to be in the null set or not in

10:50:56  24  the null set, I mean, wouldn't the person have to know

10:51:02  25  something about the case in order to make that determination?

10:51:04   1        THE WITNESS:  Of course they need to know about the

10:51:06   2   case to make the responsiveness judgments, but they don't need

10:51:10   3   to be told whether a computer system has already predicted

10:51:12   4   this document to be responsive or not responsive.  That's

10:51:16   5   something which is easy to hide from them and keep them from

10:51:20   6   being biased by.

10:51:24   7        THE COURT:  Well, I'm confused.  This is -- you

10:51:42   8   don't, but some of the people want experienced people to tell

10:51:46   9   the computer what to do.  Then the computer does it, and then

10:51:52   10  when it comes down to what the computer has done, you don't

10:51:58   11  think it has to be somebody who is knowledgeable?  I don't

10:52:02   12  know how they would know whether it was responsive or not.

10:52:04   13       THE WITNESS:  No, no, you need people who are

10:52:08   14  knowledgeable about the case to produce both the training set

10:52:12   15  and the test set, but when they are reviewing the test set,

10:52:16   16  they should not know what the computer has predicted for the

10:52:20   17  documents in the test set.  It's sort of like if you were told

10:52:26   18  that, you know, the IBM computer that won on Jeopardy has

10:52:32   19  already said this document is responsive, it would be a danger

10:52:34   20  that you would be influenced by that.

10:52:44   21       THE COURT:  I am --

10:52:44   22       THE WITNESS:  I am not speaking about you, Judge, but

10:52:46   23  a person.

10:52:48   24       THE COURT:  A person, right.

10:52:52   25       If I can't visualize something, I don't get it, and I

10:52:56  1   am having a hard time visualizing.

10:53:02  2          So you're saying they can know about the case, but

10:53:08  3   they just can't know that the computer has already said this

10:53:12  4   isn't an important document or it is an important document,

10:53:16  5   either way?

10:53:16  6          THE WITNESS:  That's correct.  So the idea would be

10:53:18  7   you would be reviewing a document on a standard review

10:53:22  8   platform, and you're simply reviewing documents, and you

10:53:24  9   simply are not told was this a document that the computer said

10:53:30  10  was a miss or was this a document that the computer said was a

10:53:34  11  hit.  You're just reviewing documents, you're bringing all

10:53:34  12  the --

10:53:34  13         THE COURT:  And this is in the test set or the null

10:53:38  14  set, in the test set?

10:53:38  15         THE WITNESS:  In the test set, regardless of whether

10:53:40  16  the test set was drawn from the hits or from the misses or

10:53:42  17  from ideally the whole population, you shouldn't know that

10:53:46  18  that's where it came from.

10:53:54  19         THE COURT:  Okay.  At least I understand that.

10:53:58  20  BY MR. McKEOWN:

10:54:00  21  Q.  And, Dr. Lewis, if one uses search terms, you still have

10:54:08  22  the possibility of -- let me back up.  Every time somebody

10:54:12  23  looks at a document in a document review, they have to mark it

10:54:16  24  responsive or nonresponsive, correct?

10:54:18  25  A.  Assuming that that's the assessment they are being asked

10:54:22　1　to make.  Sometimes people are asked to make assessments on

10:54:24　2　other kinds of topics, you know, product name, things like

10:54:28　3　that, but one of the things they might be reviewing in this

10:54:30　4　case is responsive versus nonresponsive.

10:54:34　5　Q.  And in your view -- let me ask you this.  In your view, is

10:54:36　6　it biased every time an attorney looks at a document to decide

10:54:42　7　whether or not it's responsive or not responsive?

10:54:48　8　A.  I'm not saying anything about every time attorneys look at

10:54:50　9　documents.  I'm talking about construction of a sample that

10:54:52　10　would be used to evaluate a piece of software or to evaluate

10:54:56　11　the thoroughness of a discovery process.

10:55:00　12　Q.  And your view is that regardless of who the attorney is

10:55:02　13　that is reviewing those documents, that person is not to be

10:55:06　14　trusted to make an honest decision about responsive or

10:55:10　15　nonresponsive on that null set; is that your testimony?

10:55:12　16　A.  I am not impugning anybody's honesty.  I am simply saying

10:55:18　17　it is difficult for human beings to make subjective decisions

10:55:20　18　in an unbiased fashion that are going to be used to evaluate a

10:55:24　19　piece of software when you've been told what the software

10:55:28　20　thinks the right answer is.

10:55:34　21　　　　　MR. McKEOWN:  Your Honor, at this point, I am going

10:55:36　22　to turn over the cross to Mr. Neuwirth.

10:55:38　23　　　　　THE COURT:  Do you want to take a few minutes?

10:55:42　24　　　　　MR. NEUWIRTH:  Whatever is good for the court and the

10:55:44　25　witness.

| | | |
|---|---|---|
| 10:55:44 | 1 | THE COURT: How long are you going to be, do you |
| 10:55:46 | 2 | think? Just approximately. |
| 10:55:48 | 3 | MR. NEUWIRTH: I hope to be about half an hour. |
| 10:55:50 | 4 | THE COURT: Oh, well, then why don't we take -- it's |
| 10:55:54 | 5 | 10:00 o'clock. We have been here since 8:00. So between 10 |
| 10:55:58 | 6 | and 12 minutes. Okay? Set your clocks. Okay. |
| 11:01:32 | 7 | (Short break.) |
| 11:10:26 | 8 | THE COURT: We are back on the record. |
| 11:10:32 | 9 | Mr. Neuwirth? |
| 11:10:32 | 10 | MR. NEUWIRTH: Thank you, your Honor. Good morning; |
| 11:10:34 | 11 | Steven Neuwirth, representing defendant Georgia-Pacific. |
| 11:10:40 | 12 | - - - |
| 11:10:40 | 13 | DANIEL D. LEWIS, CROSS-EXAMINATION |
| 11:10:40 | 14 | BY MR. NEUWIRTH: |
| 11:10:42 | 15 | Q. Good morning, Dr. Lewis. |
| 11:10:44 | 16 | A. Good morning. |
| 11:10:44 | 17 | Q. Now, Dr. Lewis, as we all know, you were present on the |
| 11:10:46 | 18 | first day of this hearing and you were present, correct, for, |
| 11:10:50 | 19 | among other things, the testimony of Mr. Koch from KPMG and |
| 11:10:56 | 20 | the testimony of Mr. Brown from Counsel on Call about the |
| 11:11:00 | 21 | processes that Georgia-Pacific used in this case, correct? |
| 11:11:02 | 22 | A. Yes. |
| 11:11:04 | 23 | Q. Now, do you recall that on the first day of this hearing, |
| 11:11:10 | 24 | you testified about the conclusions that you reached regarding |
| 11:11:18 | 25 | GP's, Georgia-Pacific's, ESI search methodology? |

11:11:22  1  A. Yes.

11:11:22  2  Q. And do you recall that your first conclusion was that

11:11:26  3  Georgia-Pacific's methodology can't be relied on to find a

11:11:32  4  substantial portion of responsive documents?

11:11:32  5  A. I'd have to see the transcript to know exactly the wording

11:11:36  6  I used.

11:11:36  7  Q. Let's do that.

11:11:38  8       MR. NEUWIRTH: If I may approach, your Honor?

11:11:40  9       THE COURT: Yes.

11:11:40  10      MR. NEUWIRTH: I'd like to give the witness a copy of

11:11:42  11  the transcript. Would it be convenient to your Honor to give

11:11:48  12  you one as well?

11:11:50  13      THE COURT: I think I have it, actually. What page

11:11:52  14  are you on?

11:11:56  15      MR. NEUWIRTH: Let's start, your Honor --

11:11:58  16      THE COURT: This is on direct? Was this --

11:12:00  17      MR. NEUWIRTH: Yes.

11:12:00  18  BY MR. NEUWIRTH:

11:12:00  19  Q. Let's turn to page 243 of the transcript.

11:12:04  20      THE COURT: Okay. I got it. Thank you.

11:12:04  21  BY MR. NEUWIRTH:

11:12:06  22  Q. And let's turn to line 20 and 21.

11:12:08  23      Now, I asked you, Dr. Lewis, whether your first

11:12:12  24  conclusion was that Georgia-Pacific's methodology can't be

11:12:16  25  relied upon to find a substantial portion of the responsive

11:12:20  1   documents.  So why don't you look at line 20 and 21 of page

11:12:26  2   243, which says, "My conclusions are first that it cannot be

11:12:30  3   relied upon to find a substantial portion of responsive

11:12:32  4   documents."

11:12:34  5          Does that remind you that that was your first

11:12:36  6   conclusion?

11:12:36  7   A.  Yes, that was my wording.

11:12:38  8   Q.  And you gave several reasons for this conclusion, right?

11:12:42  9   A.  Yes, I did.

11:12:42  10  Q.  And the first reason you gave was that the process of

11:12:46  11  developing queries by Georgia-Pacific was done without having

11:12:50  12  collected all the sources of responsive material, right?  That

11:12:56  13  was the first reason you gave, wasn't it?

11:12:58  14  A.  Well, let's look at the wording there.

11:13:00  15  Q.  Turn to page 244, lines 5 to 8, where you said, The first

11:13:04  16  is that the process of developing their queries was done

11:13:08  17  without having collected all of the sources of responsive

11:13:12  18  material.

11:13:12  19          So I have exactly quoted what your first reason was,

11:13:16  20  correct?

11:13:16  21  A.  That's correct.

11:13:16  22  Q.  Now, in giving that first reason, you asserted that there

11:13:22  23  are substantial sources of potentially responsive material

11:13:26  24  that are not associated with particular key custodians, right?

11:13:30  25  A.  That was my understanding from the testimony of Mr.

11:13:34 1 Hanners, who said there are often materials in an enterprise

11:13:38 2 which are not associated with particular custodians.

11:13:40 3 Q. So that is what you said at the hearing, right?

11:13:42 4 A. That's correct.

11:13:42 5 Q. You said that one of the reasons for your criticism of

11:13:46 6 Georgia-Pacific's method was that there are substantial

11:13:52 7 sources of potentially responsive material that are not

11:13:54 8 associated with particular key custodians. You said that,

11:13:54 9 right?

11:14:00 10 A. That is my understanding from Mr. Hanners is that is a

11:14:04 11 possibility.

11:14:04 12 Q. Right.

11:14:04 13 And you don't know that yourself, do you? You are

11:14:06 14 completely relying on Mr. Hanners for that point, aren't you?

11:14:08 15 A. Yes, I am.

11:14:08 16 Q. And as you sit here today, you don't have the ability, do

11:14:12 17 you, to name a single data source that Georgia-Pacific didn't

11:14:16 18 collect that you feel it should have collected, right?

11:14:18 19 A. That's correct.

11:14:20 20 Q. You just don't know, do you?

11:14:20 21 A. I do not know.

11:14:22 22 Q. And in terms of Dr. Hanners, he didn't identify a specific

11:14:32 23 data source that Georgia-Pacific didn't collect that it should

11:14:34 24 have, did he?

11:14:34 25 A. Mr. Hanners?

11:14:36  1  Q.  Mr. Hanners.

11:14:38  2  A.  I would have to go back and look at his testimony if he

11:14:40  3  described any particular computer systems.

11:14:44  4  Q.  Well, you are relying on Mr. Hanners for this point,

11:14:46  5  right?  You came here to court and said that Georgia-Pacific

11:14:52  6  hadn't done enough collection of documents.

11:14:54  7        Do you or do you not know what Mr. Hanners said about

11:14:58  8  that topic with respect to Georgia-Pacific?

11:15:00  9        MR. MOGIN:  Objection.  We are mischaracterizing the

11:15:02  10  prior testimony.

11:15:04  11        THE COURT:  Well, wait.

11:15:06  12        MR. NEUWIRTH:  I can clear this up.  I can clear this

11:15:08  13  up.

11:15:10  14        THE COURT:  I don't understand what Mr. Mogin's

11:15:14  15  objection is, though.  What's your objection?

11:15:16  16        MR. MOGIN:  Well, Dr. Lewis has testified that he

11:15:18  17  doesn't have an opinion on collection.  In fact, when

11:15:24  18  Mr. Hanners was testifying about collection, what did we have

11:15:28  19  over from this side of the aisle?  We had nothing but pop-ups

11:15:32  20  and objections.

11:15:32  21        Are we going to talk about collection in this

11:15:34  22  hearing?  Is that going to be part of it?

11:15:36  23        THE COURT:  No, I --

11:15:36  24        MR. MOGIN:  Is my witness, who said he wasn't here to

11:15:40  25  talk about collection, now going to be cross-examined about

11:15:42    1    collection?

11:15:42    2             THE COURT:  I don't think that's what Mr. Neuwirth is

11:15:46    3    doing.

11:15:46    4             MR. NEUWIRTH:  Absolutely not.

11:15:48    5             THE COURT:  He is basing from the last hearing --

11:15:50    6    Mr. Neuwirth didn't get a chance to cross-examine Dr. Lewis at

11:15:54    7    the last hearing.  And Dr. Lewis very specifically, I have

11:15:58    8    asterisked all over the place in mine, he has two opinions on

11:16:04    9    why you can't be trusted.  He is asking him about opinion

11:16:08    10   number one where he got that opinion and what he knows about

11:16:10    11   it.  Does he know anything else independently about it, I

11:16:14    12   assume.

11:16:16    13            MR. NEUWIRTH:  Correct.

11:16:16    14            MR. MOGIN:  If he wants to spend his time on this,

11:16:18    15   okay, your Honor.

11:16:18    16            THE COURT:  Well, sure.

11:16:20    17   BY MR. NEUWIRTH:

11:16:20    18   Q.  It's true, isn't it, that the first reason you gave for

11:16:22    19   criticizing Georgia-Pacific's method was to say that the

11:16:28    20   process of developing queries was done without having

11:16:30    21   collected all the sources of responsive material, right?

11:16:34    22   A.  Right.

11:16:34    23   Q.  And your sole basis for that was your reliance on

11:16:38    24   Mr. Hanners, wasn't it?

11:16:40    25   A.  Well, of course, I have personal experience in large

11:16:46  1  enterprises and know that it's possible for materials not to

11:16:50  2  be associated with particular custodians.  But my -- I am

11:16:56  3  marginally relying on Mr. Hanners that this is possible in

11:16:58  4  this case.

11:16:58  5  Q.  Largely or completely?  Are you changing your testimony

11:17:02  6  from what you said on the first day of this hearing?

11:17:04  7  A.  I am saying in the context of this case that there is the

11:17:06  8  possibility there are materials which are not associated with

11:17:10  9  a custodian, I am relying on Mr. Hanners completely.

11:17:14  10  Q.  Correct.  And what did Mr. Hanners say Georgia-Pacific had

11:17:16  11  failed to do?

11:17:16  12  A.  Well, we would have to look at Mr. Hanners' testimony.

11:17:18  13  Q.  So you came to court and expressed that opinion, but you

11:17:22  14  don't remember what Mr. Hanners said, right?

11:17:24  15        MR. MOGIN:  Objection.

11:17:24  16        THE COURT:  He can ask if he remembers what he said.

11:17:28  17  Your objection is overruled.  Okay?

11:17:30  18        THE WITNESS:  At this point, I do not remember what

11:17:34  19  Mr. Hanners' words were.

11:17:34  20  BY MR. NEUWIRTH:

11:17:34  21  Q.  So let me refresh your recollection.  I have Mr. Hanners'

11:17:38  22  report in front of me, and I will just give you examples of

11:17:40  23  what Mr. Hanners said.

11:17:42  24        One of the things he said is that there is a failure

11:17:48  25  of Georgia-Pacific to have included any former employees in

| 11:18:00 | 1 | its group of custodians.  He said, it's very unusual to find a |
| 11:18:02 | 2 | group of custodians that consists solely of current employees |
| 11:18:04 | 3 | and does not contain a single former employee. |
| 11:18:06 | 4 | Do you remember that he said that? |
| 11:18:06 | 5 | A.  I would need to look at the testimony to be sure of the |
| 11:18:10 | 6 | exact wording. |
| 11:18:12 | 7 | MR. NEUWIRTH:  May I show it to him, your Honor? |
| 11:18:12 | 8 | THE COURT:  Okay.  But I think the record should be |
| 11:18:16 | 9 | clear here.  When you're saying "his testimony," isn't that |
| 11:18:20 | 10 | his report?  I mean, just so we're clear -- |
| 11:18:26 | 11 | MR. NEUWIRTH:  Mr. Hanners' report. |
| 11:18:28 | 12 | THE COURT:  Just so we clarify, when you say |
| 11:18:30 | 13 | "testimony," most people use that that's what he testified in |
| 11:18:34 | 14 | court, and I don't remember what the heck he said. |
| 11:18:36 | 15 | MR. NEUWIRTH:  You are absolutely correct.  I am now |
| 11:18:38 | 16 | holding Mr. Hanners' report that was submitted -- |
| 11:18:42 | 17 | THE COURT:  So this is from the report, Dr. Lewis. |
| 11:18:44 | 18 | MR. NEUWIRTH:  Thank you. |
| 11:18:44 | 19 | THE WITNESS:  All right. |
| 11:18:46 | 20 | MR. NEUWIRTH:  If I may approach the witness, your |
| 11:18:48 | 21 | Honor? |
| 11:18:48 | 22 | THE COURT:  Sure.  You may. |
| 11:18:48 | 23 | BY MR. NEUWIRTH: |
| 11:18:50 | 24 | Q.  It says right here, start with, Therefore. |
| 11:18:52 | 25 | A.  Okay.  So which exhibit is this? |

11:18:58  1  Q.  That is Mr. Hanners' report.

11:18:58  2  A.  Okay.  All right.  So I have this starting at page 6 --

11:19:04  3  oh, I see.  You have kind of -- right.  So this is the cover

11:19:08  4  page.

11:19:10  5        So, therefore, it is very unusual that the final

11:19:14  6  group of custodians consists solely of current employees and

11:19:16  7  does not contain a single former employee.

11:19:20  8        Is this the thing you wanted me to read?

11:19:22  9  Q.  Right.  That's what it says, doesn't it?

11:19:24  10  A.  Yes, it does.

11:19:24  11  Q.  Now, in your preparation for giving your opinion that

11:19:30  12  Georgia-Pacific had failed to -- in your words, failed to have

11:19:36  13  collected all the sources of responsive material, did you

11:19:40  14  undertake in any way to look at the custodians, the list of

11:19:44  15  custodians that Georgia-Pacific had provided to plaintiffs

11:19:48  16  that Georgia-Pacific said it was going to be collecting

11:19:52  17  documents from?

11:19:52  18  A.  I believe I did see a list of the names.  I didn't look in

11:19:56  19  detail at it.

11:19:56  20  Q.  And do you recall looking at a letter that Britt Miller,

11:20:02  21  who is here in the courtroom today, sent to Mr. Mogin and

11:20:06  22  Mr. Freed on August 11th, 2011, that was included as Exhibit 2

11:20:12  23  to the February 6th submission to this court the defendants

11:20:16  24  made?  That was in the record before this hearing.  Did you

11:20:20  25  look at that letter?

11:20:22   1    A.  I would have to look at a copy of that again to refresh my

11:20:24   2    memory.

11:20:24   3    Q.  And did you, in preparing for the hearing, recognize that

11:20:30   4    of the custodians that were listed, four of them were

11:20:34   5    expressly identified as former employees of Georgia-Pacific?

11:20:38   6    A.  I don't recall that fact.

11:20:40   7    Q.  Now, you also assert that -- well, in making your

11:20:54   8    assertion, did you consider that Georgia-Pacific, in fact,

11:20:56   9    collected ESI that was not associated with particular

11:21:00   10    custodians?

11:21:00   11    A.  Is there a question there?

11:21:06   12    Q.  Yes.  Did you consider that in your rendering your opinion

11:21:10   13    that your first concern with Georgia-Pacific's process was

11:21:12   14    that it didn't collect the proper set of responsive documents,

11:21:16   15    potentially responsive documents?

11:21:18   16    A.  My opinion was based on Mr. Hanners' testimony that not

11:21:24   17    all sources of potentially responsive material had been

11:21:26   18    collected.

11:21:28   19    Q.  Okay.  So you have nothing to say beyond Mr. Hanners,

11:21:30   20    right, on that topic?

11:21:34   21    A.  Only based on my experience that there are sources that

11:21:36   22    are not associated with custodians.

11:21:38   23    Q.  In fact, you have made no effort yourself in rendering

11:21:42   24    this first point to really understand or evaluate exactly what

11:21:46   25    it was that Georgia-Pacific identified as the set of materials

11:21:50   1   from which it was going to do its searches?

11:21:54   2   A. Could you repeat the question?

11:21:56   3   Q. You haven't -- as you sit here today, you don't know the

11:22:02   4   full set of materials that Georgia-Pacific has collected for

11:22:06   5   purposes of finding responsive materials in this case, right?

11:22:08   6   A. That's correct.

11:22:10   7   Q. Now, the second reason you gave for saying that

11:22:14   8   Georgia-Pacific's methodology can't be relied on to find a

11:22:18   9   substantial portion of the responsive documents was that

11:22:22  10   Georgia-Pacific developed the search term queries on a sample

11:22:26  11   set of five arbitrarily selected custodians, right?

11:22:30  12   A. On a set of five arbitrarily collected custodians.

11:22:36  13   Q. Correct. You said your criticism was that

11:22:40  14   Georgia-Pacific's sample set included five arbitrarily

11:22:44  15   selected custodians, right?

11:22:46  16   A. No, I don't believe I used the word --

11:22:46  17   Q. 244, line 19 to 20: "The second reason is that

11:22:52  18   Georgia-Pacific did the development of their queries on a set

11:22:56  19   of five arbitrarily selected custodians."

11:23:00  20        Does that refresh your recollection of what you said

11:23:02  21   at the last hearing?

11:23:02  22   A. Oh, absolutely. What I was saying is I did not use the

11:23:06  23   word "sample" in that sentence.

11:23:06  24   Q. Okay. Now, you know, right, that this case concerns an

11:23:16  25   alleged conspiracy to restrict output and fix prices for

11:23:20   1   container work, correct?

11:23:22   2   A.   Yes.

11:23:22   3   Q.   And you also know that the plaintiffs are alleging that by

11:23:24   4   raising prices for container board, the defendants were able

11:23:28   5   to raise prices for boxes as well, right?

11:23:32   6   A.   That's my understanding.

11:23:34   7   Q.   So I take it in concluding that the Georgia-Pacific

11:23:38   8   custodians were arbitrarily selected, you have considered

11:23:42   9   carefully who those custodians were, right?

11:23:44   10  A.   No.

11:23:46   11  Q.   So you don't know who they were?

11:23:52   12  A.   It doesn't matter.

11:23:52   13  Q.   Oh, so your view is whoever Georgia-Pacific picked as a

11:23:58   14  custodian for the sample set, it would be arbitrary?

11:24:02   15  A.   There were, I believe, somewhere between 14 and 16

11:24:06   16  custodians who have been identified at that point.  There was

11:24:08   17  a selection of four of them plus one person who was not a

11:24:12   18  custodian who was believed to have responsive information, and

11:24:16   19  I believe that any choice of four of that set of custodians

11:24:20   20  and a choice of one person who is not believed to have

11:24:24   21  responsive information would be inappropriate.

11:24:26   22  Q.   Right.  So, now, you've said you don't know, but I can

11:24:30   23  represent to you, that one of the four custodians that was

11:24:32   24  picked for this set for testing was the senior executive who

11:24:38   25  heads GP's container board and box businesses.  Was that an

11:24:42 1  arbitrary choice?

11:24:44 2  A.  Any selection of four of the custodians that are known

11:24:50 3  from the larger set is arbitrary from the standpoint of

11:24:54 4  producing a useful sample for statistical purposes.

11:24:58 5  Q.  Right.  So in your view, picking the executive who heads

11:25:00 6  the container board and box business, picking the former

11:25:04 7  director of trades and purchasing, picking the director of

11:25:08 8  planning and operations for the container board business, and

11:25:12 9  picking the director of planning analysis with supervisory

11:25:16 10  responsibility for any decisions about when to reduce or

11:25:18 11  increase output, your view is that all of that are just

11:25:24 12  arbitrary choices, right?  Those are your words, aren't they?

11:25:26 13  A.  From a statistical standpoint, yes, those are arbitrary

11:25:30 14  choices.

11:25:30 15  Q.  In fact, you can't name a single person at Georgia-Pacific

11:25:36 16  who would have been a better choice for the sample set, can

11:25:40 17  you?

11:25:40 18  A.  It's not an issue of choosing particular people who are

11:25:44 19  better.  It's an issue of having samples that are

11:25:46 20  representative of the population of documents to be searched.

11:25:50 21  Q.  So your real point, your real point, isn't it, that

11:25:54 22  custodians shouldn't have been used at all; that instead, a

11:25:58 23  random sample should have been selected from the entire corpus

11:26:04 24  of the ESI from the relevant business, right?

11:26:08 25  A.  I am not saying that you select a sample from all the

11:26:12 1 documents in the enterprise information systems.  I am saying

11:26:14 2 you select a sample from all of the documents that have been

11:26:20 3 identified as being appropriate to search for responsive

11:26:24 4 documents.

11:26:24 5 Q.  And so you're saying that rather than use custodians for

11:26:28 6 your sample set, from whatever is the broad set of ESIs that

11:26:32 7 you have collected for the case, what you call the corpus,

11:26:36 8 from there, you should take a random sample from across that

11:26:40 9 set.  That's your point, right?

11:26:42 10 A.  Right.  The point is that you need a sample which is

11:26:44 11 representative of the body -- the entire body of material

11:26:46 12 that's going to be searched for responsive information.

11:26:48 13 Q.  And, therefore, you are saying, aren't you, that you

11:26:52 14 shouldn't use custodians to create your sample set, but that

11:26:56 15 rather, what you should do is from the body of ESI that you

11:27:00 16 have collected, select a random sample from across the entire

11:27:06 17 body of ESI?  That is what you are saying, correct?

11:27:08 18 A.  I am saying you should take a random sample from the

11:27:12 19 entire body of ESI.  There are sampling methods, such as

11:27:18 20 stratified sampling, which could make use of the fact that you

11:27:20 21 have additional information about custodians, but, again, that

11:27:24 22 would need to be done in a fashion that the sample was

11:27:26 23 representative of the entire collection to be searched.

11:27:28 24 Q.  Again, you are not changing your testimony from the first

11:27:32 25 hearing, are you, on page 247 at lines 16 to 19, where you

11:27:36　1　said, "Well, the first thing would be to identify the entire

11:27:40　2　universe of documents to which the queries are going to be

11:27:44　3　applied and then you would draw, for instance, a simple random

11:27:48　4　sample from that universe of documents"?

11:27:50　5　　　　　That is what you are saying, right, that that should

11:27:52　6　be done instead of creating a sample set and using custodians,

11:27:56　7　right?

11:27:56　8　A.　I'm not clear what you're saying about using the

11:28:00　9　custodians in the sample set.

11:28:02　10　Q.　Well, you have said that creating a sample set, as

11:28:06　11　Georgia-Pacific did, based on custodian files, you call that

11:28:12　12　arbitrary?

11:28:12　13　A.　Oh, certainly, choosing --

11:28:14　14　Q.　You testified instead, you should, "Identify the entire

11:28:18　15　universe of documents to which the queries are going to be

11:28:20　16　applied, and then you would draw, for instance, a simple

11:28:24　17　random sample from that universe of documents."

11:28:26　18　　　　　That's what you said should be done instead of what

11:28:30　19　Georgia-Pacific did, right?

11:28:30　20　A.　Right.　So I certainly said that drawing your sample only

11:28:34　21　from a subset of the custodians and leaving other custodians

11:28:38　22　out is not going to give you an appropriate statistical

11:28:42　23　estimate.

11:28:42　24　Q.　You didn't talk about custodians.　You said, Identify the

11:28:46　25　entire universe of documents to which the queries are going to

11:28:50  1  be applied, didn't you?

11:28:50  2  A.  Right.

11:28:50  3  Q.  That's what you said?

11:28:52  4  A.  Yes, I did.

11:28:52  5  Q.  Now, what have you done to investigate the size of the ESI

11:28:58  6  corpus for Georgia-Pacific's container board business?

11:29:02  7  A.  I haven't looked into that.

11:29:04  8  Q.  Right.

11:29:04  9      Now, prior to today's hearing, did you have an

11:29:08  10  opportunity to review the affidavit of Mr. Clancy from

11:29:10  11  Georgia-Pacific?

11:29:10  12  A.  I believe I looked at that briefly, but I would need to

11:29:20  13  refresh my memory --

11:29:20  14      THE COURT:  Can you tell me, is that a 30(b)(6)?

11:29:24  15  Tell me which bucket Mr. Clancy belongs in.

11:29:28  16      MR. NEUWIRTH:  This affidavit was submitted to your

11:29:30  17  Honor in the set of materials that Ms. Miller sent this past

11:29:34  18  Friday when you asked us to provide any exhibits that we may

11:29:38  19  wish to introduce at the hearing.

11:29:38  20      THE COURT:  Okay.  Who is he -- can you tell us who

11:29:42  21  Mr. Clancy is?

11:29:44  22      MR. NEUWIRTH:  Yes.  Mr. Clancy is the vice president

11:29:46  23  of information technology, or I.T., at Georgia-Pacific.  And

11:29:50  24  he has personal knowledge of the matters stated in his

11:29:56  25  affidavit, and I have an extra copy.

| | | |
|---|---|---|
| 11:29:58 | 1 | THE COURT: Mr. Mogin is standing up. |
| 11:30:00 | 2 | MR. NEUWIRTH: I am not surprised. |
| 11:30:00 | 3 | THE COURT: What do you say, Mr. Mogin? |
| 11:30:02 | 4 | MR. MOGIN: Well, the reason counsel isn't surprised |
| 11:30:04 | 5 | is because he knows how completely inappropriate this is. |
| 11:30:06 | 6 | Georgia-Pacific had its chance, they have their witnesses. |
| 11:30:10 | 7 | And now after those witnesses have testified to come forward |
| 11:30:12 | 8 | with an affidavit, to rely on the affidavit for the truth of |
| 11:30:16 | 9 | the matter in it, when I can't cross-examine the witness, and |
| 11:30:20 | 10 | now he's asking questions that assume not just the truth but |
| 11:30:26 | 11 | assume facts that haven't been admitted into evidence of this |
| 11:30:30 | 12 | witness. |
| 11:30:32 | 13 | MR. NEUWIRTH: In all fairness, your Honor -- |
| 11:30:34 | 14 | MR. MOGIN: Where are we with respect to evidence? |
| 11:30:36 | 15 | MR. NEUWIRTH: In all fairness, your Honor, we heard |
| 11:30:38 | 16 | for the first time at the last hearing Mr. Lewis' opinion that |
| 11:30:42 | 17 | this should be done based on the entire corpus of documents. |
| 11:30:46 | 18 | We just heard him testify about it, it was in the transcript, |
| 11:30:48 | 19 | and so in the context of providing material for this hearing, |
| 11:30:50 | 20 | we thought it was important to put into the record what that |
| 11:30:52 | 21 | corpus involved. |
| 11:30:54 | 22 | MR. MOGIN: More evidence -- more hearsay, your |
| 11:30:58 | 23 | Honor. More evidence that can't be tested, no |
| 11:31:02 | 24 | cross-examination. |
| 11:31:16 | 25 | THE COURT: What document? I have possible |

11:31:18  1  additional exhibits.  Can I see this?

11:31:22  2          MR. NEUWIRTH:  I believe it may be the last one.

11:31:34  3          MS. MILLER:  It's 21.

11:31:36  4          MR. NEUWIRTH:  It's 21.  Ms. Miller has reminded me.

11:31:50  5          THE COURT:  Have you read this?

11:31:50  6          MR. MOGIN:  I have.

11:31:54  7          MR. NEUWIRTH:  There can't be anything objectionable.

11:31:56  8  And, in fact, this is a cross of an expert who has expressed

11:32:00  9  an opinion.  He has now said that he never even figured out

11:32:04  10  what this information was.  We are providing that, and I'd

11:32:06  11  like to ask him some questions about it based on the opinion

11:32:08  12  that he offered at the last hearing.

11:32:28  13          THE COURT:  Well, do you agree that he could ask a

11:32:32  14  hypothetical question of an expert?

11:32:36  15          MR. MOGIN:  Of course.

11:32:36  16          THE COURT:  Of course.  So if Mr. Neuwirth were to

11:32:40  17  say, Hypothetically, if I were to tell you hypothetically that

11:32:46  18  Georgia-Pacific had da de da de da de da, how does that fit

11:32:50  19  into looking over the entire system?  I think that's what he

11:32:54  20  is saying.  And I read it for the first time.

11:32:56  21          So if he asks it hypothetically, do you still have an

11:33:04  22  objection?

11:33:04  23          MR. MOGIN:  Assuming it's properly asked as a

11:33:06  24  hypothetical, no.

11:33:06  25          THE COURT:  Can you ask it as a hypothetical?  Just

11:33:08 1 | put your facts in a hypothetical --

11:33:10 2 | MR. NEUWIRTH: Sure.

11:33:10 3 | THE COURT: -- because I think that's what you are

11:33:12 4 | asking.

11:33:16 5 | BY MR. NEUWIRTH:

11:33:18 6 | Q. Now, you testified that you didn't investigate what was

11:33:20 7 | the size of the ESI's corpus for Georgia-Pacific's container

11:33:24 8 | board business. So now I'd like you to consider the

11:33:28 9 | possibility that Georgia-Pacific has 7.6 terabytes, not

11:33:36 10 | gigabytes, terabytes of data just in the shared drives for its

11:33:40 11 | container board business as well as in headquarters' files

11:33:46 12 | related to container board and packaging. Can you consider

11:33:50 13 | that?

11:33:50 14 | THE COURT: Do you have an objection to that?

11:33:52 15 | MR. MOGIN: Your Honor, that's not a hypothetical the

11:33:54 16 | way he phrased it.

11:33:54 17 | THE COURT: Well, I think that is a hypothetical.

11:33:58 18 | Suppose they do -- there has to be a good-faith basis for

11:34:00 19 | asking a hypothetical question. Okay? I think there is a

11:34:04 20 | good-faith basis for asking that question, which then gets

11:34:10 21 | some -- because it's not going to the truth of the matter

11:34:18 22 | asserted that there are 7 million. I mean, it's not the exact

11:34:22 23 | amount that's relevant here. I don't think that's where

11:34:26 24 | Mr. Neuwirth is going, but I don't know.

11:34:28 25 | MR. NEUWIRTH: You are correct, your Honor.

| | | |
|---|---|---|
| 11:34:30 | 1 | THE COURT: So this is foundation for asking a |
| 11:34:32 | 2 | question. Objection overruled. |
| 11:34:34 | 3 | BY MR. NEUWIRTH: |
| 11:34:34 | 4 | Q. So I'd like you to assume that Georgia-Pacific has 7.6 |
| 11:34:40 | 5 | terabytes of data just in the shared drives for its container |
| 11:34:46 | 6 | board business and in its headquarters' files related to the |
| 11:34:52 | 7 | container board and box business. Okay? Just 7.6 terabytes. |
| 11:34:52 | 8 | A. Okay. |
| 11:34:56 | 9 | Q. Please assume that number. |
| 11:34:58 | 10 | A. All right. |
| 11:34:58 | 11 | Q. Now, a terabyte is 1,000 gigabytes, correct? |
| 11:35:02 | 12 | A. Correct. |
| 11:35:02 | 13 | Q. And as you went over this morning with Mr. McKeown, there |
| 11:35:10 | 14 | is a cost, isn't there, for uploading data onto a platform for |
| 11:35:14 | 15 | the purpose of applying the type of protocol that you have |
| 11:35:18 | 16 | proposed in this case? And you said, right, that different |
| 11:35:22 | 17 | vendors have different prices for doing that, correct? |
| 11:35:24 | 18 | A. Correct. |
| 11:35:24 | 19 | Q. And why don't you assume for now that that price is in the |
| 11:35:32 | 20 | range of, let's say, 400 to $800 per gigabyte. Okay? And why |
| 11:35:38 | 21 | don't we take the low end of that range, and let's say $400 |
| 11:35:44 | 22 | per gigabyte. |
| 11:35:48 | 23 | So if it's 7.6 terabytes, right, that's 7600 |
| 11:35:56 | 24 | gigabytes, correct? |
| 11:35:58 | 25 | A. Um-hmm. |

11:36:00   1   Q. And multiplying 7600 times 400 yields over 3 million,

11:36:06   2   doesn't it?

11:36:06   3   A. Yes.

11:36:08   4   Q. So if you assume that just what I have described, just the

11:36:16   5   shared servers for the container board business and the

11:36:18   6   headquarters' files for container board and boxes, if we just

11:36:22   7   assume that that limited universe is the corpus of ESI that

11:36:26   8   would be uploaded for purposes of this exercise that you have

11:36:30   9   outlined in your proposal, and if you assume that the cost is

11:36:36   10   $400 per gigabyte to do that, just to put this material on

11:36:40   11   line to run your program would cost over $3 million, right?

11:36:44   12   A. If all 7.6 terabytes of material are reasonable places to

11:36:50   13   look for responsive documents.

11:36:52   14   Q. Right. And, again, you don't know what's the reasonable

11:36:56   15   place to look; that's something that you still want everybody

11:36:58   16   else to figure out for implementing your protocol, right?

11:37:02   17   A. I have not been asked to do an analysis of what the

11:37:08   18   reasonable places to look for responsive documents are.

11:37:10   19   Q. Now, your proposal here, as we have gone over it, is not

11:37:16   20   to use the files of custodians for doing the testing, but

11:37:20   21   instead, as you put it, to take a random sample from across

11:37:26   22   the entire corpus of the ESI. We went over that several

11:37:30   23   times.

11:37:30   24   Now, having expressed that opinion here, let me ask

11:37:34   25   you if you agree with the following statement: One useful

11:37:40  1  approach is to initiate the search and retrieval process by

11:37:48  2  focusing on the information collected from a few of the

11:37:50  3  custodians who were at the center of the facts at issue in the

11:37:56  4  litigation or investigation.

11:37:58  5          Do you agree with that or not?

11:38:00  6  A.  What are you saying that those documents would be used

11:38:04  7  for?

11:38:04  8  Q.  For initiating the search and retrieval process.  It says,

11:38:10  9  One useful approach to initiate the search and retrieval

11:38:12  10  process is by focusing on the information collected from a few

11:38:18  11  of the custodians who were at the center of the facts at issue

11:38:22  12  in the litigation or investigation.

11:38:24  13         Do you agree with that or not?

11:38:24  14  A.  I'm sorry.  It says, It says?  What says that?

11:38:28  15  Q.  You want to know what that quote is from before you tell

11:38:32  16  me whether you agree with it?

11:38:32  17  A.  No, I don't need to know where the quote is from.  I am

11:38:36  18  pretty sure about that.

11:38:38  19  Q.  Do you agree with it or not?

11:38:38  20  A.  I agree that can be a reasonable strategy for obtaining

11:38:42  21  the documents to be given the training.  I would not agree

11:38:48  22  that's a reasonable strategy to produce a sample that would

11:38:50  23  let you make a valid estimate of effectiveness across the

11:38:54  24  entire population.

11:38:54  25  Q.  Do you agree with the following statement:  Focusing on

| | | |
|---|---|---|
| 11:39:00 | 1 | information collected from the core custodians which has a |
| 11:39:06 | 2 | higher likelihood of being relevant will help the team |
| 11:39:10 | 3 | efficiently develop its understanding of the issues and |
| 11:39:14 | 4 | language used by the custodians, thus allowing them to more |
| 11:39:18 | 5 | efficiently develop and implement an appropriate search and |
| 11:39:22 | 6 | retrieval process? |
| 11:39:24 | 7 | Do you agree with that? |
| 11:39:24 | 8 | A.  Again, that clearly -- that information is useful for |
| 11:39:32 | 9 | lawyers to help develop an understanding of the case, and |
| 11:39:34 | 10 | those documents would arguably be particularly useful for |
| 11:39:38 | 11 | initializing the training in the supervised learning system. |
| 11:39:42 | 12 | Q.  And this says, To develop and implement an appropriate |
| 11:39:44 | 13 | search and retrieval process, doesn't it?  That's what the |
| 11:39:48 | 14 | quote says, right, that I just read to you? |
| 11:39:50 | 15 | It doesn't limit it to these initial steps.  It says |
| 11:39:54 | 16 | that doing this would allow them to more efficiently develop |
| 11:39:56 | 17 | and implement an appropriate search and retrieval process, |
| 11:40:00 | 18 | right? |
| 11:40:00 | 19 | A.  That I -- you are quoting a document which I have not |
| 11:40:04 | 20 | seen. |
| 11:40:06 | 21 | Q.  Really?  Let me tell you what that document is.  You're |
| 11:40:08 | 22 | familiar with the Sedona Conference, aren't you? |
| 11:40:12 | 23 | A.  Yes. |
| 11:40:12 | 24 | Q.  In fact, you were just there recently? |
| 11:40:14 | 25 | A.  I was there -- |

11:40:14   1   Q.  And I take it you are familiar with the materials that the

11:40:18   2   experts in this case have submitted to the court, correct,

11:40:22   3   because you said you reviewed those materials, right?

11:40:24   4   A.  Yes.

11:40:24   5   Q.  Well, what I just read to you is from a document that's in

11:40:28   6   the record called Sedona Conference Best Practices Commentary

11:40:34   7   on the Use of Search and Information Retrieval Methods in

11:40:38   8   E-discovery at page 214.

11:40:44   9            And, in fact, in your own published writings, you

11:40:46   10  have cited to that article, haven't you?

11:40:48   11  A.  I believe I have.

11:40:50   12  Q.  Now, the third reason you gave for saying that

11:40:56   13  Georgia-Pacific's methodology can't be relied upon to find a

11:41:00   14  substantial portion of the relevant documents was your view

11:41:02   15  that spam filters were applied here, right?

11:41:08   16           There is no trick.  That's at page 245, line 19.

11:41:24   17  A.  Okay.

11:41:24   18  Q.  You said at that point that the third reason for your

11:41:28   19  criticism of the GP methodology, the Georgia-Pacific

11:41:32   20  methodology, was that spam filters were applied here, correct?

11:41:34   21  A.  What I said was the third was the question about the spam

11:41:38   22  filtering, the junk filtering that was brought up earlier.

11:41:40   23  That filtering on the string Expedia.com may have removed

11:41:44   24  responsive documents.  Now, I was a bit unclear from

11:41:48   25  Mr. Brown's testimony at what point that was no longer done.

11:41:52   1   In the November 22nd document, it was still described as part

11:41:54   2   of the process.

11:41:54   3   Q.   Now, in fact, you were present for Mr. Brown's testimony,

11:41:58   4   weren't you?

11:41:58   5   A.   Yes, I was.

11:42:00   6   Q.   And Mr. Brown testified quite unambiguously, didn't he,

11:42:04   7   that, in fact, spam filters ended up not being used here at

11:42:08   8   all, correct?

11:42:08   9   A.   Let's look at his testimony.

11:42:10  10   Q.   176, line 22, through 177, line 5.

11:42:10  11        THE COURT:   Page 17- --

11:42:10  12   BY MR. NEUWIRTH:

11:42:30  13   Q.   176, line 22, through page 177, line 5.   Page 177, line 5.

11:42:32  14   He says, And standard in every e-discovery matter that I have

11:42:36  15   ever worked in specifically deduplication and demisting, which

11:42:40  16   I believe those terms were discussed this morning, it is not

11:42:42  17   uncommon to apply a domain, a set of domain name filters that

11:42:48  18   are historically regarded in the industry as being, for lack

11:42:50  19   of a better term, junk emails or junk domain names, such as,

11:42:56  20   you know, ESPN.com, WSJ, the WallStreetJournal.com, maybe

11:43:00  21   Expedia, Orbitz, that kind of thing.   But in this particular

11:43:04  22   case, that was not done.   We have looked at that possibility,

11:43:08  23   and we just didn't find that many, many domain names of that

11:43:14  24   nature that it warranted the exercise of suppressing them.

11:43:18  25        Now, what was ambiguous about that testimony, in your

| | | |
|---|---|---|
| 11:43:22 | 1 | view? |
| 11:43:22 | 2 | A.  Well, I wasn't clear whether I should rely on the |
| 11:43:26 | 3 | testimony or on the November 22nd letter. |
| 11:43:28 | 4 | Q.  And the November 22nd letter, in your view, said that this |
| 11:43:30 | 5 | was done? |
| 11:43:30 | 6 | A.  Well, let's look at it, but I believe in the November 22nd |
| 11:43:36 | 7 | letter, it talks about Expedia.com. |
| 11:43:40 | 8 | Q.  Okay.  This is already marked as Plaintiffs' Exhibit 4. |
| 11:43:58 | 9 | MR. NEUWIRTH:  May I approach, your Honor? |
| 11:43:58 | 10 | THE COURT:  Yes. |
| 11:44:00 | 11 | From the last hearing? |
| 11:44:02 | 12 | MR. NEUWIRTH:  From the last hearing. |
| 11:44:02 | 13 | THE COURT:  I have that. |
| 11:44:04 | 14 | BY MR. NEUWIRTH: |
| 11:44:38 | 15 | Q.  Now, I take it you're referring on page 3 to the first |
| 11:44:42 | 16 | bullet point, which says that, The data set was reduced by |
| 11:44:46 | 17 | culling clear, nonresponsive junk materials by domain names, |
| 11:44:54 | 18 | such as WSJ.com, ESPN.com, Expedia.com, et cetera, right? |
| 11:44:56 | 19 | A.  Yes. |
| 11:44:58 | 20 | Q.  In fact, Mr. Brown was asked about this expressly by |
| 11:45:02 | 21 | Mr. Mogin, wasn't he, in the text that follows in the |
| 11:45:06 | 22 | transcript where we just looked, right? |
| 11:45:08 | 23 | A.  Where are you pointing? |
| 11:45:14 | 24 | Q.  Page 177 and 178.  He testified, correct, that that |
| 11:45:18 | 25 | statement in the letter was inaccurate and that, in fact, none |

| | | |
|---|---|---|
| 11:45:22 | 1 | of this domain name -- none of this domain culling was done, |
| 11:45:30 | 2 | right? |
| 11:45:30 | 3 | A.  Could you point me to exactly what lines you mean there? |
| 11:45:36 | 4 | Q.  Sure. |
| 11:45:38 | 5 | "QUESTION:  Mr. Mogin, so if you would, please refer to |
| 11:45:40 | 6 | plaintiffs' -- I'm sorry -- the November 22nd letter. |
| 11:45:42 | 7 | "ANSWER:  Thank you.  I would appreciate it if you |
| 11:45:46 | 8 | would refer to letters by their dates.  Yes. |
| 11:45:58 | 9 | A.  Shall I continue reading it for you? |
| 11:46:00 | 10 | "QUESTION:  On page 5 there, I believe you will see a |
| 11:46:02 | 11 | reference to confidence level. |
| 11:46:04 | 12 | "ANSWER:  Yes. |
| 11:46:04 | 13 | "QUESTION:  And did you author that sentence? |
| 11:46:06 | 14 | "ANSWER:  I did not author this sentence, no. |
| 11:46:08 | 15 | "QUESTION:  Just the sentence? |
| 11:46:12 | 16 | "ANSWER:  I provided information that was undoubtedly |
| 11:46:14 | 17 | used in the construction of the sentence, yes." |
| 11:46:18 | 18 | How far should I read? |
| 11:46:20 | 19 | Q.  So having read the testimony, you don't think anything |
| 11:46:22 | 20 | that Mr. Brown said in his testimony was unclear, right? |
| 11:46:26 | 21 | A.  Yes, I mean, if we have the -- well, I don't know -- I am |
| 11:46:32 | 22 | not making any statement about everything Mr. Brown said in |
| 11:46:34 | 23 | his testimony, but he does say that they did not bother to use |
| 11:46:42 | 24 | domain names. |
| 11:46:42 | 25 | Q.  Now, the next reason you gave, the fourth reason you gave |

11:46:46    1   saying that Georgia-Pacific's methodology can't be relied on

11:46:50    2   to find a substantial portion of the relevant documents, is

11:46:52    3   that during the iterative process for developing the search

11:46:56    4   terms, the reviewers determining the effectiveness of the

11:47:00    5   search terms knew whether or not the documents they were

11:47:02    6   looking at had been hit by the search terms, right?

11:47:08    7         This is the point you were talking about earlier,

11:47:10    8   right?  Correct?

11:47:10    9   A.  Yes.

11:47:10   10   Q.  And you said that that creates a bias on the part of the

11:47:14   11   reviewers, correct, because --

11:47:14   12   A.  Yes.

11:47:16   13   Q.  -- they know whether a document was in the null set or

11:47:18   14   not, correct?

11:47:18   15   A.  Yes.

11:47:18   16   Q.  Now, your proposed method that you went over in detail

11:47:22   17   with Mr. McKeown this morning calls, doesn't it, for a small

11:47:28   18   group of attorneys familiar with the details of the case to

11:47:34   19   assess the responsiveness of a sample set of documents in

11:47:36   20   order to train the software to identify responsive and

11:47:40   21   nonresponsive documents, correct?

11:47:44   22         You have a small group of attorneys that take a pile

11:47:46   23   of documents, they choose which ones are responsive and which

11:47:48   24   ones aren't, they give that information to the software, the

11:47:54   25   software then learns from that, the software then makes

11:47:56  1  choices, and then in your proposal, without knowing what the

11:48:02  2  software chose, that set of attorneys then looks at the set of

11:48:06  3  documents that was identified by the software and tells the

11:48:12  4  software which ones are responsive and which ones aren't.  And

11:48:14  5  that process continues until, as you said, the software has

11:48:18  6  learned well enough how to distinguish responsive from

11:48:24  7  nonresponsive documents, correct?

11:48:24  8  A.   Could you restate the question?

11:48:28  9  Q.   What didn't you understand about that question?

11:48:30  10  A.   Well, it was a very long description of a process and

11:48:36  11  it's --

11:48:36  12  Q.   The process that you proposed in this case, isn't it?

11:48:38  13  A.   It's the plaintiffs' process, so --

11:48:42  14  Q.   Right.  And your process calls on using a software that

11:48:46  15  has to be taught what is responsive and what isn't, correct?

11:48:52  16  A.   That is correct.

11:48:52  17  Q.   And the way it is taught is that what you have described

11:49:00  18  as a small group of attorneys that are familiar with the

11:49:02  19  details of the case look at a sample set of documents,

11:49:08  20  determine which ones they believe are responsive and which

11:49:12  21  ones they believe are not, and then those choices are told to

11:49:18  22  the software so that the software can begin to learn from

11:49:22  23  those choices, correct?

11:49:24  24  A.   That is plaintiffs' process.  I want to clarify that in

11:49:32  25  plaintiffs' process, I am not responsible for

| Time | Line | |
|---|---|---|
| 11:49:36 | 1 | characterizations of the personnel of whether they're |
| 11:49:40 | 2 | attorneys or paralegals or reviewers or whatnot. |
| 11:49:44 | 3 | Q. You said this morning that they have to be people who |
| 11:49:46 | 4 | are -- |
| 11:49:48 | 5 | A. They must be knowledgeable. |
| 11:49:48 | 6 | Q. -- familiar -- right, they must be knowledgeable about the |
| 11:49:52 | 7 | case? |
| 11:49:52 | 8 | A. Knowledgeable about the responsiveness of the documents. |
| 11:49:54 | 9 | Q. So now you are saying it could be a knowledgeable |
| 11:49:56 | 10 | paralegal, right? |
| 11:49:56 | 11 | A. I'm saying that from my standpoint as an information |
| 11:50:00 | 12 | retrieval scientist, what's important is the people be able to |
| 11:50:02 | 13 | make accurate determinations of responsiveness. I am not |
| 11:50:04 | 14 | expressing opinions about the qualifications of particular |
| 11:50:08 | 15 | types of legal personnel. |
| 11:50:10 | 16 | Q. Right. So this group of people -- |
| 11:50:10 | 17 | A. Right. |
| 11:50:10 | 18 | Q. -- within this general framework of being okay that you |
| 11:50:14 | 19 | have just described, a small number of them are going to look |
| 11:50:18 | 20 | at a set of documents, decide which ones they consider |
| 11:50:22 | 21 | responsive and which ones they don't, and give that |
| 11:50:24 | 22 | information to the software, correct? |
| 11:50:26 | 23 | A. Right. So that is for producing the training documents |
| 11:50:30 | 24 | for training the software. |
| 11:50:32 | 25 | Q. For training the software. |

11:50:34    1    And the next thing that happens is the software takes

11:50:36    2    a set of documents, and it decides based on the information it

11:50:40    3    has, which it considers responsive and which it considers

11:50:44    4    nonresponsive.  That output is then given back to the same set

11:50:48    5    of attorneys, and in your proposal, they don't know what

11:50:50    6    choices the software made, but those attorneys then pick from

11:50:54    7    that set which they consider responsive, which they do not,

11:50:58    8    and that information is then given to the software again to

11:51:02    9    learn from those choices, and in that iterative process, the

11:51:08    10   feedback from the attorneys, your proposal is to have the

11:51:10    11   software learn what is a responsive document and what isn't,

11:51:14    12   correct?

11:51:14    13   A.  No.  You're confusing two issues.

11:51:18    14       The labeling of the data for training is done by

11:51:22    15   attorneys or whoever is able to make responsible -- able to

11:51:26    16   make accurate judgments of responsiveness.  There is also data

11:51:32    17   which is a sample set to be used for evaluating the system

11:51:36    18   that also needs to be labeled by people who can make accurate

11:51:40    19   judgments of responsiveness.

11:51:42    20   Q.  Right.

11:51:42    21   A.  You are mixing the two issues together.  You implied that

11:51:46    22   the attorneys who are reviewing documents for training the

11:51:50    23   system don't know what the system's decisions were on there.

11:51:56    24   I didn't specify that they couldn't know that for the training

11:51:58    25   data.  I specified that it's important that people who are

11:52:02    1   labeling the sample to be used to evaluate the system do not

11:52:04    2   know what the system's behavior on that sample are, the test

11:52:08    3   set.

11:52:08    4   Q.  Do you agree -- you agree that there will be a small group

11:52:12    5   of attorneys that will decide which will label documents

11:52:16    6   responsive or not responsive and give that information to the

11:52:22    7   software so it can learn from those decisions how to identify

11:52:26    8   a responsive and a nonresponsive document, correct?

11:52:30    9   A.  I am not characterizing the type of personnel.

11:52:32    10  Q.  You agree that there are going to be a small number of

11:52:34    11  human beings that do that, correct?

11:52:36    12  A.  That seems likely that it will be a small number, yes.

11:52:40    13  Q.  Right.  Because that's what you recommend, isn't it?

11:52:42    14  A.  Yes, indeed.

11:52:42    15  Q.  Now, that happens at the start of the process in one of

11:52:50    16  your early steps, correct?

11:52:52    17  A.  So if we look at the protocol --

11:52:56    18  Q.  I am just asking if it happens in one of your early steps.

11:52:58    19  You know that the answer to that is yes, don't you?

11:53:02    20  A.  Yes, yes.

11:53:02    21  Q.  So the answer to that is yes.  Let's just focus on the

11:53:06    22  questions being asked.

11:53:08    23          If the answer to that is yes, you agree, don't you,

11:53:12    24  that whatever reasons those human beings may have for making

11:53:20    25  the choices they make will be taught to the software, correct?

11:53:26    1    Whatever decisions those human beings make are what the

11:53:30    2    software will learn, right?

11:53:30    3    A.   The software learns from the example documents that have

11:53:40    4    been assessed as responsive or nonresponsive.  The software is

11:53:44    5    obviously not aware of the reasons that are in people's heads.

11:53:46    6    Q.   But whatever biases those human beings may bring to their

11:53:50    7    selection process will be taught to the software, correct?

11:53:54    8    A.   That's a complicated issue in the science of machine

11:54:02    9    learning.  The --

11:54:04    10   Q.   Whatever biases they have will be taught, won't it, to the

11:54:08    11   software?

11:54:08    12          MR. MOGIN:  Your Honor, is it possible that the

11:54:10    13   witness could be allowed to finish his answers before we move

11:54:12    14   to the next question?

11:54:14    15          THE COURT:  You're asking me to rewrite all of

11:54:22    16   cross-examination.

11:54:24    17          MR. MOGIN:  No, I am really not.  I am just asking

11:54:26    18   for a simple courtesy here.

11:54:28    19          THE COURT:  Well, okay.

11:54:32    20          Carolyn, would you read back the question for me?

11:54:34    21     (Record read.)

11:54:46    22          THE COURT:  So let's let Dr. Lewis answer that.

11:54:50    23          MR. NEUWIRTH:  Yes.

11:54:50    24          THE WITNESS:  So I am a little unclear when you're

11:54:54    25   using the word "biases" if you are talking about biased in the

11:55:02  1  cyclical sense or if you're talking about some tendency doing

11:55:06  2  the review to make responsiveness decisions one way or the

11:55:10  3  other.

11:55:10  4  BY MR. NEUWIRTH:

11:55:12  5  Q.  Well, if the people doing the review have a belief, if

11:55:18  6  those human beings have a belief, that certain types of

11:55:22  7  documents shouldn't be included as responsive, that would be

11:55:26  8  taught to the software, correct?

11:55:26  9  A.  Yes, to the extent to which the software is able to

11:55:32  10  accurately model those decisions, it will attempt to do so.

11:55:38  11  Q.  You have a question of whether the software can accurately

11:55:40  12  model those decisions?  I thought the whole point of your

11:55:44  13  process here was to use the software because it can accurately

11:55:50  14  model those decisions?

11:55:50  15  A.  The point is the supervised learning software is -- in my

11:55:52  16  opinion, will likely give you the best chance of modeling

11:55:56  17  those decisions.

11:55:56  18  Q.  But you don't know if it can?

11:55:58  19  A.  I mean, it's very easy to pose tech classification

11:56:02  20  problems that no computer could solve.  I mean, I would love

11:56:08  21  to pose the tech classification system that told me the

11:56:08  22  newspaper stories about companies whose stock is going to go

11:56:14  23  up next week, but no computer is going to be able to do that

11:56:16  24  very well.

11:56:16  25  Q.  Well, in this case, this case is about whether the

11:56:18  1   defendants allegedly purposely reduced the output of their

11:56:22  2   production as a means to raise prices and did it through

11:56:28  3   collusion.  If one of the reviewing -- if one of the reviewers

11:56:30  4   in this initial set that's used to teach the software decides

11:56:36  5   that certain types of communications should not be included as

11:56:40  6   responsive, the software will learn it, won't it, not to

11:56:46  7   include those types of communications, right?

11:56:48  8   A.  Right.  So if all of the reviewers who are labeling

11:56:52  9   documents for training are consistent on that point, certain

11:56:58  10  types of communications are not considered responsive, the

11:57:00  11  system will, you know, attempt to build a model that will

11:57:10  12  reproduce that.

11:57:10  13       Now, if they have some insight into the

11:57:12  14  communication, for instance, that is not in the text or the

11:57:14  15  metadata, obviously the system can't -- can only look at

11:57:18  16  what's there.  But it's going to do its best job to reproduce

11:57:24  17  the distinctions that the reviewers are making.

11:57:26  18  Q.  But you don't disagree, do you, that in your method, you

11:57:34  19  propose to have a small number of people be involved in giving

11:57:38  20  this guidance to the software for it to learn, right?  You

11:57:42  21  said that.

11:57:42  22  A.  Yes.

11:57:42  23  Q.  You said a small number of senior attorneys.  Today, you

11:57:46  24  are saying it might be a paralegal.  But you said a small

11:57:48  25  number of people.  That's in your proposal, isn't it?

11:57:50  1  A.  Plaintiffs' proposal refers to concepts like attorneys and

11:57:58  2  paralegals and whatnot.  I am not responsible for the

11:58:00  3  characterization of particular types of personnel.

11:58:04  4  Q.  So you're changing your testimony again?

11:58:06  5  A.  No, I am not.

11:58:06  6  Q.  Please look at the last day of testimony, page 258,

11:58:10  7  line 22.

11:58:16  8       Now, you were shown Plaintiffs' Exhibit 10, which

11:58:18  9  Mr. McKeown went over with you this morning, which is

11:58:22  10  plaintiffs' proposed CBA search process for each defendant.

11:58:24  11  And you were asked by Mr. Mogin, the plaintiffs' attorney:

11:58:28  12       "QUESTION:  Dr. Lewis, is this the proposed process

11:58:30  13  that you have put together?

11:58:32  14       "ANSWER:  Yes.  I aided the plaintiffs in the

11:58:36  15  development of this, and in particular, I am responsible for

11:58:40  16  the technical aspects of this proposal."

11:58:44  17       So is this your proposal, or now are you just going

11:58:46  18  to call it the plaintiffs' proposal?

11:58:48  19  A.  It is the plaintiffs' proposal.  I am responsible for the

11:58:52  20  technical aspects of this proposal.  I am not responsible for

11:58:56  21  the characterization of the particular types of legal

11:59:00  22  personnel who carry out the proposal.

11:59:02  23  Q.  So this idea of having a small number of attorneys who are

11:59:08  24  familiar with the case teach the software what is responsive

11:59:12  25  and not responsive, is that something the plaintiffs came up

11:59:14   1    with, or is that something you came up with?

11:59:16   2    A. I would, in my professional opinion, say that it should be

11:59:22   3    people who understand the case and can make accurate

11:59:26   4    determinations of responsiveness.

11:59:28   5    Q. I didn't ask you what your opinion is. I asked you who

11:59:32   6    came up with this idea that's in here? Are you going to

11:59:34   7    attribute that to the plaintiffs, or are you going to take

11:59:36   8    responsibility for it?

11:59:36   9    A. I am responsible for the characterization of the people as

11:59:42   10    being able to make accurate responsiveness judgments. I am

11:59:46   11    responsible, in my opinion, that it is easier with a small

11:59:52   12    number of personnel to achieve consistency in response to

11:59:56   13    those judgments. I am not expressing an opinion about what

12:00:00   14    their status in the legal world is.

12:00:02   15    Q. Right. And that's my point, what you just said. The

12:00:04   16    smaller the number of people, as you just said, the more

12:00:08   17    likely there is to be consistency in their judgments, right?

12:00:12   18    You just said that, right?

12:00:12   19    A. It is easier to achieve consistency among a smaller group.

12:00:16   20    Q. Right. So if there is a bias about excluding certain

12:00:20   21    documents that that small group of people has, that will be

12:00:26   22    taught to the software, won't it? Because you just said there

12:00:28   23    is consistency, and the whole point here is to have the

12:00:30   24    software learn from what those people teach it, right? Your

12:00:36   25    method depends on that happening, doesn't it?

12:00:38   1   A.   Could you restate the question?

12:00:40   2   Q.   You really didn't understand that?

12:00:42   3   A.   Well, there was, I believe, at least two separate

12:00:44   4   questions in there.

12:00:46   5   Q.   You agree, don't you, that the smaller the number of

12:00:54   6   reviewers, the more likely you are to have consistency in the

12:00:58   7   decisions they make about what to tell the software as

12:01:02   8   responsive and not a responsive document, right?  You said

12:01:06   9   that several times today, right?

12:01:08  10   A.   It's easier to achieve consistency with a small team of

12:01:12  11   people.  Obviously, there's factors such as the training the

12:01:14  12   people have and the processes they go through to compare

12:01:18  13   answers, for instance, and things like that.

12:01:20  14   Q.   There are all sorts of variables, we understand.  There

12:01:22  15   are all sorts of variables that can determine how effective

12:01:24  16   this training of the software is going to be.

12:01:26  17   A.   Right.

12:01:26  18   Q.   And none of those variables or what to do about that is

12:01:32  19   detailed in your proposal, right?  But that has to be dealt

12:01:34  20   with, right?

12:01:36  21   A.   There are process details that have to be dealt with in

12:01:40  22   carrying out any information retrieval exercise.

12:01:42  23   Q.   Including yours?

12:01:42  24   A.   Including mine.  The process is important in any exercise.

12:01:46  25   Q.   So now we know you have a -- a couple of reviewers, a

12:01:52   1   small number of reviewers who are going to teach the software

12:01:54   2   what it has to consider responsive or nonresponsive, whatever

12:02:00   3   variables have to be dealt with presumably would be dealt with

12:02:02   4   one way or another, and now don't you agree that whatever

12:02:06   5   views those small number of people bring to the process and

12:02:10   6   what they consider to be responsive or not responsive will be

12:02:14   7   taught to the software?

12:02:14   8   A.   Right.   Whatever responsiveness assessments they make of

12:02:20   9   the training data are what the software is going to attempt to

12:02:24  10   model.

12:02:24  11   Q.   And so if they have a bias -- for example, Mr. Mogin says

12:02:28  12   that the case should cover a certain topic but these reviewers

12:02:32  13   don't agree with that -- they can teach the software not to

12:02:34  14   pick that up, right?

12:02:36  15   A.   So now you're talking about a situation where you have

12:02:42  16   multiple reviewers but they're systematically disagreeing?

12:02:46  17   Q.   Well, you said that a small number of reviewers are the

12:02:48  18   most likely to agree with each other to be consistent?

12:02:52  19   A.   I said it's easiest to achieve consistency in training a

12:02:56  20   small set of reviewers.

12:02:58  21   Q.   Right.

12:03:00  22        Now, when you criticized what Georgia-Pacific did

12:03:06  23   here, you said this morning that you were concerned that if

12:03:12  24   reviewers like Mr. Brown knew that the documents he was

12:03:18  25   looking at came from the null set, he would presumably be

12:03:24  1  biased towards finding that those documents were not

12:03:28  2  responsive?  That's the bias you're worried about, right?

12:03:32  3  A.  That is a concern that one has to have whenever you are

12:03:36  4  evaluating the sample that's going to be used to -- whenever

12:03:42  5  you're -- excuse me -- assessing a sample is going to be used

12:03:46  6  to evaluate a system, it is a potential source of bias if the

12:03:50  7  person doing that assessment knows what the behavior of the

12:03:52  8  system was.

12:03:54  9  Q.  Now, isn't it equally possible that knowing documents come

12:03:58  10  from the null set, Mr. Brown and the others who were involved

12:04:04  11  could be biased towards finding that documents are responsive?

12:04:08  12  Isn't that an equal possibility, that they could bend over

12:04:10  13  backwards to say something is responsive rather than, in your

12:04:14  14  view, to say it's not?  Aren't those both equal possibilities?

12:04:16  15  A.  Yes, they are.

12:04:18  16  Q.  And, in fact, didn't Mr. Brown testify here that what he

12:04:24  17  did was when he looked in the null set to see what was

12:04:28  18  responsive, he bent over backwards to include even documents

12:04:34  19  that he considered marginally responsive?  Meaning even when

12:04:36  20  he wasn't asserting they were responsive, he was counting them

12:04:40  21  as responsive.  He said that, didn't he?

12:04:42  22  A.  Yes, he did, and that was a form of bias.

12:04:44  23  Q.  Right.  But that's a bias that helps the plaintiffs

12:04:48  24  because it means he's finding more things to be responsive

12:04:50  25  rather than less, which is the bias you say you're worried

12:04:54   1   about?

12:04:54   2   A.  My goal is not to help one side or the other.  My goal is

12:04:58   3   to try to design a process that would yield a statistically

12:05:02   4   accurate estimate of the recall of the system.  I am not here

12:05:04   5   for one side or the other to win.

12:05:06   6   Q.  And, in any event, you're not an expert on psychology,

12:05:14   7   right?

12:05:14   8   A.  No, I would not characterize myself as an expert on

12:05:18   9   psychology.  I have some exposure to psychological principles

12:05:22   10   that are relevant to the design of information retrieval

12:05:26   11   systems.

12:05:26   12   Q.  Right.  But you don't consider yourself an expert on which

12:05:30   13   bias would apply in looking at a null set, the bias to bend

12:05:32   14   over backwards to include things or the bias that you say

12:05:36   15   might be there to exclude things?  You have no professional

12:05:38   16   way to know which is more likely, do you?

12:05:40   17   A.  I have only the common sense -- you know, sense of that

12:05:46   18   people, if they are aware that a system that they are involved

12:05:52   19   in using is going to be evaluated in some fashion, I have only

12:05:56   20   the common sense that they may have an interest in that system

12:06:04   21   looking good.

12:06:04   22   Q.  Right.  And then you also have -- apart from what you call

12:06:06   23   your common sense general knowledge, you also have the

12:06:10   24   specific testimony here from Mr. Brown that he and his team

12:06:12   25   bent over backwards to include documents that they even

12:06:16  1  considered marginally responsive, right?

12:06:18  2  A.  I am aware that he has, in fact, testified to that.

12:06:28  3  Q.  Now, you also said that the review of documents in the

12:06:30  4  null set has to be consistent with the review of documents

12:06:34  5  that's done for the actual production, right?  You want the

12:06:36  6  review of the sample to be based on the same principles as the

12:06:40  7  review that is done for production, right?

12:06:44  8  A.  Could you point to the particular place you are talking

12:06:48  9  about there?

12:06:50  10  Q.  250, line 18 to 21.

12:07:02  11        THE COURT:  This is Dr. Lewis' testimony --

12:07:04  12        MR. NEUWIRTH:  Dr. Lewis' testimony in his direct by

12:07:06  13  Mr. Mogin.

12:07:06  14        THE COURT:  250, what did you say?

12:07:08  15        MR. NEUWIRTH:  250; page 250, let's start at line 10.

12:07:22  16  BY MR. NEUWIRTH:

12:07:22  17  Q.  This is Dr. Lewis:  My understanding of Mr. Brown's

12:07:24  18  testimony of the description in the November 22nd letter, to

12:07:26  19  the extent that I could make out what they are saying in that

12:07:28  20  letter, is that their validation process is focused on

12:07:32  21  determining how many responsive documents are in the null set.

12:07:36  22  And that can be a reasonable way to evaluate an information

12:07:40  23  retrieval system.  There are some caveats that are important,

12:07:44  24  however.

12:07:46  25        Question by Mr. Mogin:

| | | |
|---|---|---|
| 12:07:46 | 1 | "QUESTION:  What are the caveats? |
| 12:07:48 | 2 | "ANSWER:  Number one, well, the caveats are that it's |
| 12:07:54 | 3 | extremely important if you do that that the review of the |
| 12:07:58 | 4 | documents in the null set be consistent with the review of the |
| 12:08:00 | 5 | documents that's done for production." |
| 12:08:04 | 6 | That's exactly what you said at the last hearing, |
| 12:08:06 | 7 | right? |
| 12:08:06 | 8 | A.  That's correct. |
| 12:08:06 | 9 | Q.  You are not changing that testimony today? |
| 12:08:08 | 10 | A.  No, I am not. |
| 12:08:10 | 11 | Q.  And then the reason you said that that was not done here, |
| 12:08:16 | 12 | right, was you said that Counsel on Call only had three people |
| 12:08:20 | 13 | reviewing the test set while a larger group of 15 or more were |
| 12:08:26 | 14 | reviewing the documents from the entire corpus hit by the |
| 12:08:30 | 15 | search terms to determine their responsiveness for production |
| 12:08:34 | 16 | to plaintiffs, right? |
| 12:08:34 | 17 | A.  Well, I would prefer to read what my actual testimony was |
| 12:08:40 | 18 | here. |
| 12:08:40 | 19 | Q.  Okay.  Go to line 23.  You were asked:  Did |
| 12:08:44 | 20 | Georgia-Pacific follow the process; namely, that the review of |
| 12:08:48 | 21 | documents in the null set be consistent with the review of |
| 12:08:50 | 22 | documents that's done for production?  Mr. Mogin asked you if |
| 12:08:54 | 23 | Georgia-Pacific followed that process. |
| 12:08:56 | 24 | And your answer was, Well, I am not aware of who did |
| 12:08:58 | 25 | the review for production.  Well, no, there is a whole -- |

| | | |
|---|---|---|
| 12:09:02 | 1 | that's right, because Mr. Brown testified to that.  Mr. Brown |
| 12:09:04 | 2 | and two of his colleagues reviewed the null sets or samples, |
| 12:09:10 | 3 | excuse me, samples from the null set, and I think there was a |
| 12:09:12 | 4 | team of I think 15 lawyers or something that were doing the |
| 12:09:16 | 5 | review for production.  So what would be critical is to, you |
| 12:09:20 | 6 | know, have some statistical guarantee that the two sets of |
| 12:09:24 | 7 | reviewers are actually making comparable decisions.  The |
| 12:09:28 | 8 | serious problem that could arise is that if you use a |
| 12:09:30 | 9 | stringent criterion for evaluating responsiveness in the null |
| 12:09:34 | 10 | set and then a more liberal criterion during the review and |
| 12:09:38 | 11 | then actually compared the quantities, you could think, well, |
| 12:09:40 | 12 | you found some, you know, incredibly large portion of the |
| 12:09:42 | 13 | documents that were out there, but that ratio would not be |
| 12:09:46 | 14 | correct because they were not being reviewed competently. |
| 12:09:52 | 15 | Do you see that? |
| 12:09:52 | 16 | A.  Yes. |
| 12:09:52 | 17 | Q.  And so if you expressed a concern, right, that because |
| 12:09:54 | 18 | there were three lawyers looking at the null set and 15 |
| 12:09:58 | 19 | lawyers reviewing documents that had been hit by the search |
| 12:10:04 | 20 | terms to determine whether they should be produced to the |
| 12:10:06 | 21 | plaintiffs, you saw a potential problem of consistency, right? |
| 12:10:10 | 22 | A.  Yes. |
| 12:10:14 | 23 | Q.  Now, your proposal or what you're now sometimes calling |
| 12:10:18 | 24 | the plaintiffs' proposal, whoever it comes from, the proposal |
| 12:10:20 | 25 | that's been given to the court expressly says, doesn't it, |

| | | |
|---|---|---|
| 12:10:24 | 1 | that the teaching to the software should be done by a small |
| 12:10:30 | 2 | number of people, right? And that makes sense because you |
| 12:10:36 | 3 | told us today that's a good way to guarantee consistency in |
| 12:10:38 | 4 | what's taught to the software, right? |
| 12:10:40 | 5 | A. Yes. |
| 12:10:40 | 6 | Q. Now, in a case where there's already millions of pages of |
| 12:10:46 | 7 | documents, just doing what the defendants have done, and there |
| 12:10:52 | 8 | could be tens of millions of documents added if we were |
| 12:10:56 | 9 | ordered to do what you and the plaintiffs or you or the |
| 12:10:58 | 10 | plaintiffs or whoever it is is suggesting that we do, is it |
| 12:11:04 | 11 | really your view that it's wrong for Georgia-Pacific to have a |
| 12:11:10 | 12 | team of 15 people assigned to reviewing those documents? Do |
| 12:11:14 | 13 | you really think they should limit that to just three people |
| 12:11:16 | 14 | so we can wait maybe five years to get the documents reviewed? |
| 12:11:20 | 15 | A. No, and I did not suggest that. |
| 12:11:24 | 16 | Q. Well, you said there is a problem that there were only |
| 12:11:26 | 17 | three people reviewing the documents in the null set as |
| 12:11:32 | 18 | compared to 15 who were reviewing the documents for |
| 12:11:34 | 19 | production. So which should happen? Should the number of |
| 12:11:38 | 20 | people reviewing the null set go up, or should the number of |
| 12:11:40 | 21 | people reviewing documents for production go down? |
| 12:11:42 | 22 | A. If we are using the indirect -- the so-called indirect |
| 12:11:50 | 23 | method -- sorry, the only thing that's important here is that |
| 12:11:58 | 24 | the sample that's being used to evaluate be reviewed with the |
| 12:12:04 | 25 | -- consistently between the two. Now, I guess there's several |

12:12:08   1   things going on here, but you don't have to use -- the central

12:12:14   2   thing, which I think to answer your question, is you don't

12:12:16   3   have to use all of the data that's been reviewed for training.

12:12:22   4          So, you see, you might have a large staff of

12:12:24   5   reviewers who are producing the production set.  That doesn't

12:12:28   6   mean that all of the responsive judgments that they make have

12:12:30   7   to be fed back in training.  You could have the judgments that

12:12:34   8   are made only by a smaller group of people if that was

12:12:36   9   desirable.

12:12:38  10   Q.  So, in fact, the thing that you criticized Georgia-Pacific

12:12:40  11   for on the first day, having a small number of people review

12:12:44  12   the null set and a larger number of people review the

12:12:46  13   documents for production, actually makes perfect sense and

12:12:50  14   it's the only realistic way to get the document production

12:12:54  15   done, isn't it?

12:12:54  16   A.  What I was criticizing the Georgia-Pacific process was the

12:12:58  17   potential for inconsistent judgments in the sample that's used

12:13:04  18   to evaluate the system.

12:13:04  19   Q.  Okay.  So this line that we just read, And there was a

12:13:08  20   team of I think 15 lawyers or something that were doing the

12:13:10  21   review for production compared to what you said were Mr. Brown

12:13:14  22   and two of his colleagues reviewing the null sets, that whole

12:13:18  23   testimony you gave last time we should just throw out because

12:13:20  24   it's not critical to what you are saying, right?

12:13:22  25   A.  The testimony here was relevant to the use of a sample --

12:13:30  1 | samples in evaluating the effectiveness of the Boolean queries

12:13:36  2 | that Georgia-Pacific produced.

12:13:36  3 | Q.  And they used three people to review the null set and 15

12:13:38  4 | to review the documents for production.  So is that a problem

12:13:42  5 | or is it not, the fact that they used three for one and 15 for

12:13:46  6 | the other?

12:13:46  7 | A.  It is a potential problem if the three people who are

12:13:50  8 | reviewing a sample from the null set do not have consistent

12:13:56  9 | responsiveness judgments with the 15 people.  And, in fact,

12:14:00 10 | it's not clear to me how many of those 15 people were involved

12:14:02 11 | in reviewing samples from the hit set.  That's not made clear.

12:14:06 12 | Q.  But this problem, this hypothetical problem you are

12:14:10 13 | talking about, and it is purely hypothetical, right, because

12:14:14 14 | you don't know of any difference in what they do, you are just

12:14:16 15 | speculating that maybe there is a difference, isn't that

12:14:18 16 | hypothetical problem going to exist in any process where you

12:14:22 17 | use a small number of people to do one thing and a large

12:14:26 18 | number of people to do another?  It's just inevitable, isn't

12:14:30 19 | it, that what you said before, that a small number of people

12:14:32 20 | are more likely to be consistent than a larger group, that's

12:14:36 21 | true in any process where you use different groups of people

12:14:40 22 | to do different things, right?

12:14:40 23 | A.  But there is no obligation to use different groups of

12:14:44 24 | people to review a modest-sized random sample which is used

12:14:48 25 | purely for the purpose of evaluating the system.

12:14:50  1   Q.  But you're criticizing Georgia-Pacific for using 15 people

12:14:54  2   to review documents for production to the plaintiffs?

12:14:56  3   A.  I am not criticizing that.

12:14:58  4   Q.  Okay.  You said -- we just read your testimony, and that's

12:15:00  5   why I said, Can we now throw this testimony out.  When you

12:15:04  6   were asked whether Georgia-Pacific followed the process of

12:15:06  7   having consistency between the review of the null sets and the

12:15:10  8   review of the documents for production, your answer was, I am

12:15:14  9   not aware of who did the review for production.  We will know

12:15:18  10  there is a hole that's right because Mr. Brown testified about

12:15:22  11  that.  Mr. Brown and two of his colleagues reviewed the null

12:15:24  12  sets of the samples -- excuse me, samples from the null sets,

12:15:28  13  and there was a team of I think 15 lawyers or something that

12:15:30  14  was doing the review for production.  So what would be

12:15:34  15  critical is to, you know, have some statistical guarantee that

12:15:36  16  the two sets of reviewers are actually making comparable

12:15:40  17  decisions.

12:15:40  18        Now, you talked expressly about the difference in the

12:15:44  19  numbers, and you said you want some guarantee that they're

12:15:46  20  making the same decisions.  That's not how human beings work,

12:15:52  21  but we are dealing with the real world of having to review a

12:15:56  22  very large corpus of documents and one that will be

12:16:00  23  megagigantic if you get what you're proposing here.

12:16:02  24        Don't we have to have a larger group of people

12:16:06  25  reviewing the documents for production than we do working on

12:16:08  1  the sample set or teaching the software what to do?  Or do you

12:16:08  2  not care about those practicalities and just want perfection

12:16:12  3  in the system at a statistical level?

12:16:16  4  A.  Could you ask the question again?

12:16:18  5  Q.  And what about that -- what about that don't you

12:16:20  6  understand, Dr. Lewis?

12:16:22  7  A.  It was a very long statement.

12:16:24  8  Q.  Well, you're a statistician.

12:16:26  9       MR. MOGIN:  Objection.  It was a compound question.

12:16:30  10  Can we do it that way?

12:16:30  11       THE COURT:  I agree with that.  If you want to break

12:16:36  12  it down, ask shorter questions to it, you may do that.

12:16:36  13  BY MR. NEUWIRTH:

12:16:40  14  Q.  Let me just ask one short question.

12:16:44  15       Do you agree or disagree that in the practical, real

12:16:48  16  world of document production, it makes sense to have a larger

12:16:56  17  group of people review documents for production than the

12:17:02  18  number of people that will be involved in reviewing sample

12:17:06  19  sets or teaching information to software?

12:17:10  20  A.  I agree with that.

12:17:12  21  Q.  So then the fact that Georgia-Pacific had 15 people

12:17:16  22  reviewing documents for production and only three reviewing

12:17:20  23  the null set makes perfect sense in the real world, doesn't

12:17:26  24  it?

12:17:28  25       You just said you agree it makes sense to have more?

12:17:30　1　A.　I agree using a large set of people for production may, in

12:17:34　2　fact, be appropriate in large-scale cases.

12:17:36　3　　　　What I am saying is important is that if a null set

12:17:40　4　testing approach is -- if a process is used which is drawing

12:17:46　5　samples from the null set and comparing that to the responsive

12:17:54　6　documents that are found in hit sets, it's important that

12:17:58　7　those two quantities be consistent.

12:18:00　8　Q.　And you have no basis other than speculation at this point

12:18:04　9　to say that they weren't consistent, right?

12:18:06　10　A.　I have only the evidence of potential sources of bias here

12:18:12　11　and the information retrieval literature on the --

12:18:16　12　Q.　Right, from the literature, from the academic literature;

12:18:18　13　but you have nothing in this case that would allow you to

12:18:20　14　conclude that there was any discrepancy, right?

12:18:22　15　A.　Yes, that's correct.  There was no information that was

12:18:26　16　provided me on any index or consistency studies or any tests

12:18:28　17　that would have shown the degree of consistency among

12:18:32　18　reviewers.  We know absolutely nothing about the consistency

12:18:36　19　of the reviewers that Georgia-Pacific used.

12:18:36　20　Q.　Right.

12:18:38　21　　　　Now, your second conclusion in criticizing the

12:18:42　22　methodology that Georgia-Pacific used was you said that

12:18:44　23　Georgia-Pacific cannot produce a statistically valid estimate

12:18:50　24　of effectiveness, right?

12:18:52　25　A.　Let's look at that line.

12:18:56    1   Q. 252, line 6 through 10.

12:19:06    2   A. Okay.

12:19:06    3   Q. Quote -- this is your counsel, Mr. Mogin, or plaintiffs'

12:19:14    4   counsel, Mr. Mogin: "Very good. Now, let's go to your other

12:19:18    5   area of criticism, if you will, which was that the methodology

12:19:22    6   cannot be relied upon to produce a statistically valid

12:19:26    7   estimate of effectiveness."

12:19:28    8         Your answer: "That's correct."

12:19:30    9         So that's your second criticism, right?

12:19:32   10   A. That's correct.

12:19:32   11   Q. Now, you have asserted, haven't you, that in order for

12:19:38   12   there to be value reporting, you need to have a confidence

12:19:42   13   level, what you went over with Mr. McKeown this morning; a

12:19:48   14   margin of error, which you went over with Mr. McKeown this

12:19:50   15   morning; and then the actual error rate, which you went over

12:19:54   16   with Mr. McKeown this morning, right? Those would be three

12:19:58   17   things you would need to have to be able to have valid

12:20:02   18   reporting, right?

12:20:04   19         THE COURT: What's the third?

12:20:04   20         MR. NEUWIRTH: The actual error rate.

12:20:10   21         THE WITNESS: Assuming that one wants to express a

12:20:14   22   confidence interval on error rate, the error rate is the

12:20:18   23   effectiveness measure that you are using, then you would need

12:20:22   24   your estimate of error rate, your central value, and your

12:20:28   25   margin of error, and your confidence level.

12:20:28  1  BY MR. NEUWIRTH:

12:20:30  2  Q.  Right.  So you need what you went over with Mr. McKeown

12:20:32  3  this morning.  You need to know what is the percentage

12:20:36  4  likelihood that this can be replicated over the overall

12:20:38  5  sample, whether it's 99 percent, 95 percent, or whatever the

12:20:42  6  number you're aiming for, correct?

12:20:44  7  A.  And one of the components is the confidence level which

12:20:48  8  expresses how likely you would be to see that result over a

12:20:50  9  large number of random samples.

12:20:52  10  Q.  You need your error rate, which is plus or minus 5

12:20:54  11  percent, or whatever percent you're using, right?

12:20:58  12  A.  Actually, that's the margin of error.

12:21:00  13  Q.  Margin of error, I'm sorry, plus or minus 5 percent margin

12:21:04  14  of error, right?

12:21:04  15        And then you need the actual error rate:  I found in

12:21:06  16  the null set that 3 percent of the documents actually were

12:21:10  17  responsive.  That's your error rate, your actual error rate,

12:21:12  18  right?

12:21:12  19  A.  That would be the error rate within the null set.

12:21:18  20  Q.  Now, Georgia-Pacific, in fact, has provided all of this

12:21:20  21  information, hasn't it?  It's provided the size of the null

12:21:26  22  set that was tested, right?  It provided that was 63,211

12:21:32  23  documents, right?

12:21:32  24  A.  Could you point to the document that that's reported in?

12:21:36  25  Q.  Sure.  I will show you the latest version of that.  Let me

| 12:21:38 | 1 | mark as Plaintiffs' Exhibit -- I'm sorry -- Defendants' |
| 12:21:42 | 2 | Exhibit 10. |
| 12:21:46 | 3 | MR. MOGIN: May we have a date of this version? |
| 12:21:48 | 4 | MR. NEUWIRTH: As soon as I get to it. |
| 12:21:50 | 5 | It's the March 21st, 2012, letter to Mr. Mogin from |
| 12:21:54 | 6 | me that was discussed earlier today. |
| 12:22:00 | 7 | May I approach, your Honor? |
| 12:22:00 | 8 | THE COURT: Yes. |
| 12:22:06 | 9 | MR. NEUWIRTH: I have extra copies for you, your |
| 12:22:34 | 10 | Honor, if you need an extra copy. |
| 12:22:34 | 11 | May I approach, your Honor? |
| 12:22:34 | 12 | THE COURT: Yes. |
| 12:22:36 | 13 | Was that in the big book? |
| 12:22:38 | 14 | MR. NEUWIRTH: Yes. |
| 12:22:40 | 15 | THE COURT: What number is the big book? |
| 12:22:44 | 16 | MS. MILLER: I believe it's 11. |
| 12:22:44 | 17 | MR. NEUWIRTH: Thank you, Ms. Miller. |
| 12:22:46 | 18 | THE COURT: So this, now, is defendants' what? |
| 12:22:48 | 19 | MR. NEUWIRTH: I marked it as Defendants' Exhibit 10, |
| 12:22:50 | 20 | but if it's easier for the court, we can just treat it as the |
| 12:22:52 | 21 | exhibit there. Why don't we mark it as Exhibit 10 so it's in |
| 12:22:56 | 22 | the record. |
| 12:22:56 | 23 | THE COURT: I think you should. |
| 12:23:00 | 24 | BY MR. NEUWIRTH: |
| 12:23:00 | 25 | Q. Now, you reviewed this document before you testified |

12:23:02  1  today, right?

12:23:02  2  A.  Yes, I did.

12:23:02  3  Q.  So this document points out, doesn't it, at page 3 of 4,

12:23:10  4  that the size of the null set for final testing was 63,211

12:23:16  5  documents, right?

12:23:16  6  A.  Where on the page are you?

12:23:20  7  Q.  At the carryover from page 3 to 4.  If you look at the

12:23:26  8  very bottom.

12:23:28  9  A.  Okay.

12:23:28  10  Q.  It says that the 83,544 documents comprising the testing

12:23:34  11  corpus for the October 2011 validation, 20,333 documents were

12:23:40  12  hit by the search terms, leaving the remaining 63,211

12:23:46  13  documents as the null set?

12:23:46  14  A.  I see that.

12:23:48  15  Q.  Okay.  We also know, don't we, how the random sample from

12:23:54  16  this null set was created, right?  Mr. Brown testified that he

12:24:02  17  used the Raosoft software to create the random sample, right?

12:24:08  18  A.  Could we look at Mr. Brown's testimony?

12:24:14  19  Q.  Sure, page 180, lines 3 to 7.

12:24:28  20  A.  Okay.

12:24:30  21  Q.  Right.

12:24:30  22      So this was a software tool called Raosoft, right?

12:24:32  23  A.  I see that --

12:24:34  24  Q.  So Raosoft was used to figure out the sample set size that

12:24:38  25  was needed, right?

12:24:40    1    A.  I see that.

12:24:40    2    Q.  And then we had testimony from Mr. Koch from KPMG that

12:24:50    3    KPMG employed the randomizing function Clearwell to pull

12:24:54    4    random samples based on the sample size that Raosoft had

12:24:56    5    called for, right?

12:24:58    6    A.  I would have to look at Mr. Koch's testimony.

12:25:00    7    Q.  Okay.  Turn to page 64, line 12.

12:25:06    8         You were here for that testimony, right?

12:25:08    9    A.  Yes, I was.

12:25:08    10   Q.  64, line 12.  We applied and pulled the random sample

12:25:16    11   based on the information we were given so they would say

12:25:18    12   please create a random sample of this many documents and we

12:25:24    13   would provide that in Clearwell and then provide the results.

12:25:28    14        So you created the random sample?

12:25:30    15        "ANSWER:  Technically, we clicked the button in

12:25:34    16   Clearwell to create the random sample."

12:25:36    17        Right?

12:25:36    18   A.  I see that.

12:25:36    19   Q.  Right.  So you were there when this was described.

12:25:40    20        Now, we know, don't we, that the sample size of the

12:25:46    21   null set documents was 660 documents, right, which is what the

12:25:54    22   Raosoft program had said was needed to get a 99 percent

12:26:00    23   confidence level, right?

12:26:00    24   A.  Well, let's look at the testimony about --

12:26:06    25   Q.  Well, let's first look at the March 21st letter on page 3.

| | | |
|---|---|---|
| 12:26:22 | 1 | Mr. Brown testified that 27 -- I'm sorry. 660 |
| 12:26:28 | 2 | documents in the sample set are referenced there. Do you see |
| 12:26:34 | 3 | that in the paragraph request 1-E? |
| 12:26:38 | 4 | A. Sorry. Where are you looking again? |
| 12:26:38 | 5 | Q. On page 2 of the March 21st letter. The reference is |
| 12:26:46 | 6 | to -- |
| 12:26:46 | 7 | A. 660 documents in the sample set, correct? |
| 12:26:50 | 8 | Q. And, I'm sorry, page 3. I apologize. I apologize. |
| 12:26:52 | 9 | Page 3. |
| 12:26:52 | 10 | In request 1-J, talking about the last validation. |
| 12:26:58 | 11 | A. Okay. |
| 12:26:58 | 12 | Q. It refers to the 660 documents that were pulled, right? |
| 12:27:04 | 13 | A. So I see down here specifically 28 documents, footnote 2, |
| 12:27:12 | 14 | 4.2 percent of the 660 documents randomly selected to comprise |
| 12:27:20 | 15 | the statistically significant null set, which was based on a |
| 12:27:20 | 16 | 99 percent confidence level with a margin of error of plus or |
| 12:27:24 | 17 | minus 5 percent would be determined to be arbitrarily |
| 12:27:28 | 18 | responsive to plaintiffs' document request. |
| 12:27:30 | 19 | Q. Right. |
| 12:27:30 | 20 | So using the Raosoft program to determine the number |
| 12:27:32 | 21 | of documents that were needed to get a 99 percent confidence |
| 12:27:36 | 22 | level with a plus or minus 5 percent margin of error, 660 |
| 12:27:42 | 23 | documents were pulled from the null set of 63,211 documents. |
| 12:27:52 | 24 | And then for your error rate of 4.2 percent, we have that from |
| 12:27:58 | 25 | having found 28 documents which were determined to be at least |

| | | |
|---|---|---|
| 12:28:02 | 1 | arguably responsive to plaintiffs' document request, right? |
| 12:28:06 | 2 | A.  I see that. |
| 12:28:06 | 3 | Q.  So you have all the information you need here for a |
| 12:28:10 | 4 | statistical report, right?  You have the number of documents |
| 12:28:12 | 5 | that were in the null set, you have the sample set created |
| 12:28:16 | 6 | just as you described to Mr. McKeown this morning so that you |
| 12:28:20 | 7 | could have a certain confidence level here, 99 percent, with a |
| 12:28:26 | 8 | margin of error of plus or minus 5 percent, and you have the |
| 12:28:28 | 9 | result? |
| 12:28:30 | 10 | A.  Yes, in this March 21 letter, all three of those values |
| 12:28:34 | 11 | are specified. |
| 12:28:34 | 12 | Q.  Right.  And, in fact, it's true, isn't it, that a lot of |
| 12:28:40 | 13 | this information was in prior submissions to the court, wasn't |
| 12:28:42 | 14 | it, or you don't know that? |
| 12:28:46 | 15 | A.  You would have to point me to the prior submissions. |
| 12:28:48 | 16 | Q.  But do you know when you gave your testimony, when you |
| 12:28:50 | 17 | gave your testimony at the last hearing, that this method that |
| 12:28:54 | 18 | Georgia-Pacific used somehow wasn't -- let me use your exact |
| 12:29:06 | 19 | words -- that it somehow wasn't -- that it wasn't able to |
| 12:29:26 | 20 | produce a statistically valid estimate of effectiveness. |
| 12:29:30 | 21 | Had you determined how much of this information was |
| 12:29:32 | 22 | already in the record, or did you just make that assertion |
| 12:29:34 | 23 | without figuring it out? |
| 12:29:36 | 24 | A.  Oh, I looked very carefully, and I was not able to find |
| 12:29:40 | 25 | the actual central value of the confidence interval. |

12:29:42  1   Q.  Even though a lot of it was in the testimony earlier that

12:29:46  2   day from Mr. Brown, right?

12:29:48  3   A.  I don't recall if Mr. Brown testified to the particular

12:29:52  4   information that was admitted in the November 22nd letter.

12:29:56  5   Q.  Let me ask you, are you familiar with someone named Ralph

12:30:02  6   Losey, L-o-s-e-y?

12:30:04  7   A.  Yes, I am.

12:30:04  8   Q.  And is he an expert in any field?

12:30:08  9   A.  He is a lawyer who has considerable prominence in the area

12:30:14  10  of electronic discovery.

12:30:16  11  Q.  And you noted during your testimony, didn't you, that even

12:30:26  12  with your proposed method, it's going to be important to

12:30:32  13  consult with whatever vendor is selected to figure out what

12:30:36  14  that vendor's best practices are for creating a sample set,

12:30:42  15  correct?  That's what you told Mr. McKeown this morning,

12:30:46  16  right?

12:30:46  17  A.  Yes.

12:30:46  18  Q.  And here, Georgia-Pacific did exactly that with two

12:30:48  19  vendors, KPMG and Counsel on Call, correct?

12:30:52  20  A.  Would you point me to where that was described?

12:31:02  21  Q.  You were here for the first day of their testimony,

12:31:04  22  correct, and you heard about what their involvement was in the

12:31:10  23  creation of the sample set, right?

12:31:12  24  A.  I would like to see where -- I mean, I did hear their

12:31:16  25  testimony, but to any particular detail, I would have to go

12:31:18    1    over that again.

12:31:20    2    Q. Okay.

12:31:20    3    A. And I also would like to stress that while one should

12:31:28    4    always consult with vendors about the appropriate way to use

12:31:32    5    their software, whether for sampling or training or any other

12:31:36    6    purpose, there are basic principles of statistics which need

12:31:44    7    to be followed in order to achieve statistically valid

12:31:46    8    results.

12:31:46    9    Q. And what vendors follow those statistical methods at the

12:31:52    10    present time?

12:31:52    11    A. I would have to look at -- I mean, I have not analyzed,

12:32:00    12    for instance, the random sampling functions in particular

12:32:04    13    e-discovery platforms --

12:32:04    14    Q. So as you sit here today, you couldn't even tell

12:32:06    15    defendants which vendors they should use for this program you

12:32:10    16    provided because you don't know which vendors use the

12:32:14    17    statistical methods that you say are important to consider,

12:32:16    18    right?

12:32:16    19    A. The relevant aspects of vendor selection for the

12:32:20    20    e-discovery software that are crucial are laid out at the end

12:32:26    21    of the protocol --

12:32:26    22    Q. Well, you have a set of --

12:32:28    23    A. -- of plaintiffs --

12:32:28    24    Q. You have a set of ways that someone can pick a vendor, but

12:32:32    25    you can't tell us which vendor to use, right, because you

12:32:36    1    don't even know a vendor that does what you are asking us to

12:32:40    2    do, do you?

12:32:40    3    A.   I have looked at the supervised learning capabilities that

12:32:44    4    have been described in the literature by several vendors, so I

12:32:48    5    do know that there are several vendors who could carry out

12:32:52    6    this process.

12:32:52    7    Q.   Yes, in literature, academic literature, promotional

12:32:56    8    literature; but you have never worked with any of these

12:32:58    9    vendors to actually do this in a case like this, have you?

12:33:02   10    A.   I have analyzed data sets that were collected during

12:33:08   11    e-discovery processes as part of algorithm development for

12:33:12   12    Kroll Ontrack.

12:33:12   13    Q.   I didn't ask what you did to develop algorithms.  I asked

12:33:16   14    whether you worked with any of these vendors to do the type of

12:33:20   15    thing that you put in this document or the plaintiffs put in

12:33:22   16    this document that you want the court to order the defendants

12:33:24   17    to do instead of what they have already done?

12:33:26   18    A.   I am not asking the court to do anything.  I am giving my

12:33:32   19    analysis of the processes.  I have not worked as a consultant

12:33:38   20    on active e-discovery matters using these technologies.

12:33:42   21    Q.   But you have been out, haven't you, as an advocate for

12:33:46   22    these technologies, right?  You have given presentations in

12:33:50   23    many forums in recent years advocating a move towards this

12:33:54   24    type of machine learning, correct?

12:33:56   25    A.   I have given a number of presentations in which I

12:34:02    1    expressed my belief that supervised learning such as rank

12:34:06    2    retrieval can bring additional efficiencies to the e-discovery

12:34:12    3    process.

12:34:12    4    Q. That it can. You have been an advocate for that. And

12:34:16    5    this case, this protocol that's been put together for the

12:34:20    6    court to order the defendants to implement, this is an

12:34:24    7    opportunity to have what you have been advocating finally

12:34:28    8    required in a case, isn't it?

12:34:30    9    A. I am always interested in people using the best possible

12:34:34   10    information retrieval technology. I have worked for 25 years

12:34:38   11    in a variety of fields to teach and inform people about the

12:34:46   12    power of various sorts of information retrieval technology.

12:34:48   13    Q. And notwithstanding all those speeches and articles and

12:34:52   14    academic forum and other things you've done, you can't even

12:34:58   15    tell us a vendor that does it?

12:35:00   16    A. Is there a question there?

12:35:04   17    Q. You can say yes or no or whatever else you want to say.

12:35:10   18    A. Can you repeat the question?

12:35:12   19    Q. Notwithstanding all that you have done to promote this

12:35:18   20    type of approach, notwithstanding the articles you've read and

12:35:24   21    advocacy that you've done on behalf of this approach, you

12:35:30   22    can't even tell the defendants in this case what vendor is

12:35:32   23    able to do the things that either you or the plaintiffs are

12:35:36   24    proposing in that protocol that the court is being asked to

12:35:42   25    order the defendants to implement, right?

12:35:42  1  A.  I can tell you vendors that can implement this protocol.

12:35:46  2  Q.  Oh, so your testimony before, you're changing now?

12:35:48  3  A.  No, I am not changing my testimony.

12:35:54  4         MR. NEUWIRTH:  No further questions.

12:35:54  5         THE COURT:  Thank you.

12:36:00  6         So do you want to do redirect?

12:36:12  7         MR. MOGIN:  It might be more efficient to wait until

12:36:16  8  after Mr. Regard's testimony, your Honor, as you previously

12:36:18  9  suggested.

12:36:28  10         THE COURT:  Well, you could do redirect on what we

12:36:30  11  heard this morning because what I was saying about calling --

12:36:34  12  giving Dr. Lewis an opportunity to comment on what Mr. Regard

12:36:42  13  is going to say, because he wouldn't have heard him already.

12:36:44  14  That's what I was talking about.  But you could do redirect

12:36:48  15  right now and then we could break for lunch.

12:36:52  16         MR. MOGIN:  Okay.

12:36:52  17         THE COURT:  You seem to have some points you wanted

12:36:54  18  to make or clarify, and I am following along pretty carefully

12:37:00  19  here, and I think this is supposed to be educating Chris and

12:37:10  20  I.  So that's what this exercise is about.

12:37:12  21                          - - -

12:37:12  22         DANIEL D. LEWIS, REDIRECT EXAMINATION

12:37:12  23  BY MR. MOGIN:

12:37:12  24  Q.  Dr. Lewis, let's start off with the question you weren't

12:37:14  25  allowed to answer.  What vendors could implement?

| | | |
|---|---|---|
| 12:37:16 | 1 | THE COURT:  Well, do you really need to -- okay. |
| 12:37:24 | 2 | MR. MOGIN:  Would a sidebar be helpful, your Honor? |
| 12:37:26 | 3 | THE COURT:  Yeah, I do, actually. |
| 12:37:26 | 4 | (Whereupon, a discussion was had at the bench outside the |
| 12:39:28 | 5 | hearing of the court reporter.) |
| 12:39:28 | 6 | THE COURT:  We are back on the record.  Do you want |
| 12:39:36 | 7 | to ask your next question? |
| 12:39:38 | 8 | MR. MOGIN:  Certainly. |
| 12:39:40 | 9 | BY MR. MOGIN: |
| 12:39:44 | 10 | Q.  Dr. Lewis, counsel asked you about some information that |
| 12:39:52 | 11 | was contained in the March 21st letter, which they have marked |
| 12:39:56 | 12 | as Exhibit 3. |
| 12:39:58 | 13 | A.  Okay. |
| 12:39:58 | 14 | Q.  In particular, the information which has no citation |
| 12:40:08 | 15 | regarding the number of documents in the test that hits and |
| 12:40:16 | 16 | the null set.  Do you see that? |
| 12:40:16 | 17 | A.  I see that. |
| 12:40:18 | 18 | Q.  That's in the information that's in the carryover from |
| 12:40:20 | 19 | pages 3 to 4? |
| 12:40:22 | 20 | A.  Yes. |
| 12:40:22 | 21 | Q.  Did you have all of that information at the time that you |
| 12:40:24 | 22 | were rendering your opinion last month? |
| 12:40:28 | 23 | A.  No. |
| 12:40:28 | 24 | Q.  But you relied on a letter, did you not, from -- also |
| 12:40:36 | 25 | signed by Mr. Neuwirth as part of the factual basis for your |

12:40:40  1  testimony, correct?

12:40:40  2  A.  Yes.

12:40:42  3  Q.  Is that the same letter that Mr. Neuwirth is now

12:40:46  4  criticizing you for relying on, as opposed to the testimony of

12:40:50  5  Mr. Brown?

12:40:52  6  A.  It's the November 22nd letter which Mr. Neuwirth was

12:41:00  7  discussing.  I guess I wouldn't characterize it as

12:41:04  8  criticizing.

12:41:06  9  Q.  So if there was a discrepancy between what Mr. Brown

12:41:08  10  testified to and Mr. Neuwirth attempted to exploit that

12:41:12  11  discrepancy as between the November 22nd letter, wouldn't you

12:41:16  12  be between the proverbial rock and a hard place as to whether

12:41:20  13  to rely on the testimony of a witness or the letter from

12:41:24  14  counsel?

12:41:24  15  A.  Well, it did certainly set up a bit of a conflict for me.

12:41:32  16  I am not a lawyer, so I am not so sure if one of those should

12:41:34  17  be taken to trump the other.

12:41:36  18  Q.  All right.  Dr. Lewis, under the terms of your proposal,

12:42:02  19  isn't it, in fact, designed to allow most companies with some

12:42:04  20  supervised learning capability to be used?

12:42:08  21  A.  That's the attempt in designing the technical aspects of

12:42:12  22  the proposal was that if a company had a platform that

12:42:16  23  supports supervised learning and that it would be able to be

12:42:20  24  used for this purpose.

12:42:22  25  Q.  So you're not in any way, shape, or form attempting to

12:42:26  1  drive the business to particular vendors?

12:42:30  2  A.  No, not at all.

12:42:30  3  Q.  You are not suggesting particular vendors?

12:42:34  4  A.  No.

12:42:34  5  Q.  The key is not as -- is simply that you can show

12:42:40  6  statistically valid testing protocol?

12:42:42  7  A.  Right.  So the idea is that the protocol describes how to

12:42:48  8  use statistical rank retrieval and supervised learning and

12:42:50  9  also how to use statistical methods to evaluate the

12:42:54  10  effectiveness of those.  Indeed, the statistical evaluation

12:43:02  11  could be done even if the platform didn't support that.  I

12:43:06  12  mean, one could do the random sampling in some other fashion,

12:43:10  13  one could use an online calculator in the way that defendants

12:43:12  14  have.

12:43:12  15      And so -- in fact, I did not include anything about

12:43:18  16  the statistical calculations like computing confidence

12:43:24  17  intervals in the proposals because you could do that in some

12:43:28  18  other software.

12:43:28  19  Q.  Now, one of the principles of science, is it not, is to

12:43:32  20  try to remove the possibility of human bias or human error

12:43:38  21  affecting testing?  Is that correct?

12:43:42  22  A.  Whenever possible.

12:43:42  23  Q.  And, for example, it's common, is it not, in the medical

12:43:46  24  profession when new drugs are being tested that certain people

12:43:52  25  in a sample set or in a population will be given placebos,

12:43:56    1   correct?

12:43:56    2   A.  Correct.

12:43:56    3   Q.  And others will be given the drug that's being tested,

12:44:00    4   correct?

12:44:00    5   A.  Correct.

12:44:00    6   Q.  And those that are the patients, as it were, don't know

12:44:08    7   whether they are getting the placebo or getting the drug,

12:44:10    8   correct?

12:44:10    9   A.  Yes.  Indeed, in what's called a double blind experiment,

12:44:14    10  neither the patient nor the doctor knows which the patient has

12:44:20    11  gotten, and only third party scientists then analyze the

12:44:24    12  results afterwards.  So even the doctor -- there is the

12:44:28    13  attempt to make sure even the doctor is not biased by their

12:44:30    14  knowledge.

12:44:32    15  Q.  And is that a sort of double blind methodology to remove

12:44:36    16  bias used in other forms of statistical studies?

12:44:40    17  A.  Oh, yes.  Anywhere where there is a degree of human

12:44:46    18  subjectivity in interpreting results of the experiment, it's

12:44:50    19  desirable to do that.  I mean, obviously, there can be

12:44:54    20  situations where it can't be done, but whenever possible,

12:44:56    21  that's done.

12:44:58    22  Q.  Now, you were asked a series of questions about confidence

12:45:04    23  levels, correct?

12:45:04    24  A.  Yes.

12:45:04    25  Q.  Is it proper to compute the confidence level based on one

| | | |
|---|---|---|
| 12:45:10 | 1 | random sample in a corpus the sizes that we're talking about |
| 12:45:14 | 2 | here? |
| 12:45:14 | 3 | A. Oh, sure. You can draw a single random sample and compute |
| 12:45:22 | 4 | confidence interval with a certain confidence level. The |
| 12:45:24 | 5 | confidence level will depend on the size of the random sample, |
| 12:45:24 | 6 | and the margin of error will depend on the size of the random |
| 12:45:28 | 7 | sample. |
| 12:45:28 | 8 | Q. Do you know how many random samples have been done on any |
| 12:45:30 | 9 | of the defendants other than Georgia-Pacific's documents? |
| 12:45:34 | 10 | A. How many random samples have been done? |
| 12:45:36 | 11 | Q. Yes. |
| 12:45:38 | 12 | A. I don't know that with any certainty. There was a chart |
| 12:45:42 | 13 | that suggested that at least one random sample had been done, |
| 12:45:46 | 14 | but I have not seen other material describing that. |
| 12:45:48 | 15 | Q. Now, you testified previously, did you not, that recall is |
| 12:45:52 | 16 | perhaps one of the most important measurements of |
| 12:45:56 | 17 | effectiveness of a search technique, correct? |
| 12:45:58 | 18 | A. Yes. Particularly in an e-discovery context where there |
| 12:46:02 | 19 | is a premium on finding a large proportion of the responsive |
| 12:46:06 | 20 | documents, recall is a natural way to measure that. |
| 12:46:10 | 21 | Q. Okay. So give us the five-second definition of recall, if |
| 12:46:12 | 22 | you would, please. |
| 12:46:14 | 23 | A. Recall is of all the material that's been agreed upon in |
| 12:46:20 | 24 | each search, it is the proportion of responsive documents that |
| 12:46:24 | 25 | have been found out of all of the responsive material in the |

12:46:32    1   collection.  So it's what percentage of the good stuff did you

12:46:34    2   find.

12:46:34    3   Q.  Okay.  And have you been provided with the information

12:46:38    4   that you would need in order to compute recall for any of the

12:46:42    5   defendants?

12:46:42    6   A.  None of the other six defendants at the table, no.  One

12:46:52    7   could --

12:46:52    8   Q.  You're referring to Mr. Regard's table?

12:46:56    9   A.  Yes, with Regard's table, there is nothing there that

12:46:58    10  would let you compute recall.

12:47:00    11          In the November 22nd letter from Georgia-Pacific, if

12:47:04    12  one ignores all the problems with the sampling, the bias, and

12:47:08    13  things like that, I believe one could extract a recall figure

12:47:14    14  from that, but I would not be confident that that would be a

12:47:18    15  valid estimate.

12:47:18    16  Q.  Why not?

12:47:20    17  A.  Well, because of the bias in the review and -- oh, sorry.

12:47:26    18  I should mention, it would only be a measure of the recall

12:47:30    19  with respect to those four custodians.  We would have no

12:47:34    20  information about the recall with respect to the entire

12:47:36    21  Georgia-Pacific collection.

12:47:38    22  Q.  Okay.  Now, you testified earlier that there is a

12:47:44    23  distinction between a confidence level and a probability.  Do

12:47:50    24  you recall that testimony?

12:47:50    25  A.  Yes.

12:47:50  1  Q.  Would you explain that, please.

12:47:52  2  A.  Okay.  Well, that's sort of one of the hardest things to

12:47:58  3  explain about frequentist statistics.  So the confidence level

12:48:00  4  for a confidence interval refers to the fact that -- it's --

12:48:06  5  basically, it is a way to characterize the process by which

12:48:10  6  the confidence interval is produced.  So that if you repeated

12:48:16  7  that process a large number of times, a 95 percent confidence

12:48:22  8  level indicates that 95 percent of the time you repeated that

12:48:26  9  sampling process, you would get a confidence interval that

12:48:30  10  contained the true value.

12:48:32  11  Q.  So the confidence interval is a statistic that tells us

12:48:36  12  about the sampling process, correct?

12:48:40  13  A.  Yes.

12:48:40  14  Q.  It's not -- so if someone has a 95 percent confidence

12:48:46  15  level, that doesn't necessarily mean, does it, that the

12:48:48  16  probability is that 95 percent of the documents within the

12:48:54  17  margin of error have been retrieved by the system?

12:48:56  18  A.  Oh, not at all.  No, no, it's only a description of the

12:49:00  19  sampling process.  So, for instance, you could have 95 percent

12:49:04  20  confidence that the recall was, you know, 10 percent plus or

12:49:08  21  minus 5 percent, so you would be 95 percent confident that

12:49:12  22  your recall was between 5 and 15 percent, that you had only

12:49:16  23  found between 5 and 15 percent of the responsive material.

12:49:20  24  Q.  By the way, Dr. Lewis, how many documents does it take to

12:49:28  25  prove a price fixing conspiracy?

| | | |
|---|---|---|
| 12:49:32 | 1 | A.  Well, I'm not a lawyer, but my understanding is that, you |
| 12:49:42 | 2 | know, a single document could be used to -- a document that on |
| 12:49:44 | 3 | the face of it said, Let's fix prices or something would be |
| 12:49:48 | 4 | sufficient.  I mean, that's only my layperson's understanding. |
| 12:49:50 | 5 | Q.  Okay.  And if that document said, They are with us, if |
| 12:49:56 | 6 | that turned out to be the key document, which method is more |
| 12:50:00 | 7 | likely to find that document, Boolean searching or supervised |
| 12:50:08 | 8 | learning as you proposed it? |
| 12:50:08 | 9 | A.  Well, supervised learning would be more apt to find it |
| 12:50:12 | 10 | unless somebody had created a Boolean query that was |
| 12:50:16 | 11 | specifically designed to retrieve any known responsive |
| 12:50:20 | 12 | document. |
| 12:50:20 | 13 | But if you don't know, I mean, the point is you don't |
| 12:50:22 | 14 | know whether it might be they are with us, or it might be any |
| 12:50:28 | 15 | other unbelievable number of other ways to express |
| 12:50:30 | 16 | disagreement, what the supervised learning can do is take |
| 12:50:34 | 17 | advantage of the statistical properties of metadata as well as |
| 12:50:36 | 18 | the words, and it's likely to have the best chance of finding |
| 12:50:40 | 19 | a very difficult document like that. |
| 12:50:42 | 20 | Q.  Now, you mentioned about a query that was customized, I |
| 12:50:52 | 21 | think you said in your testimony, in order to find a |
| 12:50:52 | 22 | particular document? |
| 12:50:54 | 23 | A.  Oh, sorry.  You were talking about my testimony about |
| 12:50:58 | 24 | Georgia-Pacific or about the example I just gave? |
| 12:51:00 | 25 | Q.  The example you just gave. |

12:51:02  1    A.  Well, what I was saying, you know, you could write a

12:51:08  2    Boolean -- if you knew through magic that there was an

12:51:12  3    important document out there that said, They are with us,

12:51:16  4    there is a way to express with proximity operators in Boolean

12:51:20  5    queries or even just quotation marks to sometimes support it

12:51:24  6    that you could retrieve all the documents that had exactly

12:51:28  7    that phrase, but you would have to know in advance that that

12:51:30  8    was exactly like a code phrase or something that was being

12:51:32  9    used.

12:51:34  10            So if you don't have that knowledge in advance, the

12:51:36  11   Boolean query is not going to be very helpful for finding

12:51:40  12   documents where all the terms are high frequency.

12:51:44  13            THE COURT:  Hold on, though.  But it is possible with

12:51:50  14   the Boolean search to find documents that can have -- I mean,

12:52:02  15   if you went back over 10 years of antitrust cases and looked

12:52:06  16   at the intent, looked at the kinds of documents that were

12:52:10  17   introduced, the kind of evidence that they introduced, and you

12:52:12  18   wanted to do a Boolean word search with some of the kind of

12:52:18  19   evidence that they used in other cases, that would be

12:52:24  20   possible, correct?

12:52:24  21            These are cases that went to trial, there was

12:52:26  22   evidence, some of the evidence would have letters in it, some

12:52:30  23   of them, the more recent ones, might have e-mail, they may

12:52:34  24   have documents, they could have something with the Boolean

12:52:38  25   search, you could -- it is possible to create a document that

12:52:44  1  could capture some of that, correct?

12:52:48  2  THE WITNESS:  One can certainly use Boolean queries

12:52:52  3  to try to find documents on particular topics.  It would be --

12:52:56  4  I think the difficulty is that if somebody was trying to be

12:53:00  5  coded or indirect in their language, there is an almost

12:53:04  6  infinite variety of ways to do that.  So that even if one

12:53:08  7  looked at a huge backlog of antitrust, and even if there was

12:53:12  8  examples of people using all sorts of coded language, you

12:53:16  9  wouldn't expect that the next use of coded language would be

12:53:20  10  something that you can anticipate from that.

12:53:32  11  THE COURT:  Even with your system -- not your system,

12:53:36  12  but the CBAA system, you are still telling the machine to do

12:53:44  13  something.  Now, I understand that you're telling them to go

12:53:48  14  over departments, you are telling them to go over groups of

12:53:52  15  words, you are telling them, but you're still telling the

12:53:56  16  machine.  The machine on its own isn't coming up with, I am

12:54:00  17  with you.  I mean, you're getting input into the machine of

12:54:04  18  what the machine is looking for, right?

12:54:08  19  THE WITNESS:  Right.  So the machine would be

12:54:12  20  learning from examples of responsive documents.

12:54:14  21  THE COURT:  So you have already found a respons- --

12:54:20  22  so somehow you have a responsive document in your hand, and

12:54:24  23  you're then telling the machine to go find more that are like

12:54:28  24  that, right?

12:54:32  25  THE WITNESS:  Right.  And where like that can include

12:54:34   1   not only aspects of the textual data but aspects of the

12:54:38   2   metadata, where a file was stored, who the custodian was, what

12:54:42   3   point in time, what time of day it was sent by.  There's a

12:54:46   4   variety of --

12:54:46   5        THE COURT:  But in Boolean searches, they can also do

12:54:52   6   metadata.  I mean, that kind of --

12:54:54   7        THE WITNESS:  That's right.  Many Boolean searches

12:54:56   8   also allow search on metadata.

12:54:58   9        THE COURT:  Okay.  That's a needle in a haystack.

12:55:04  10        Back to the -- this is called I want the smoking gun.

12:55:10  11   So go back to I want the smoking gun.

12:55:14  12   BY MR. MOGIN:

12:55:14  13   Q.  Is there a difference in the metadata that would be used

12:55:16  14   in a Boolean search methodology as opposed to the metadata

12:55:20  15   that would be used in supervised learning?  In other words, do

12:55:24  16   they both use metadata in the same fashion?

12:55:28  17   A.  They could both use the same metadata fields, but they

12:55:34  18   would be using them in different fashions.  So the supervised

12:55:36  19   learning system could differentially weight the metadata in a

12:55:44  20   fashion which would not be possible to express with a Boolean

12:55:48  21   query.

12:55:48  22   Q.  And what would be the effect of the differential weight?

12:55:50  23   A.  Well, the issue is that if one is looking for documents

12:56:00  24   which are -- basically, if the evidence that something is

12:56:08  25   responsive is that there are certain patterns in the metadata.

12:56:12   1   So, for instance, that, you know, documents related --

12:56:16   2   previous responsive documents have been found, for instance,

12:56:20   3   tend to be sent in the middle of the night and that this was

12:56:24   4   unusual in comparison to nonresponsive documents or even that

12:56:30   5   they were sent from a particular email account or that

12:56:34   6   afterwards, they were stored in a particular portion of the

12:56:36   7   file system, a variety of sort of things like that, these

12:56:40   8   could be arguably very difficult to anticipate in writing a

12:56:46   9   Boolean query because there's so many different possible forms

12:56:50   10  of that evidence.  But these are the kind of pattern finding a

12:56:54   11  supervised learning method can do.

12:56:56   12  Q.  So is it correct then that in order to find what her Honor

12:57:02   13  just described as a smoking gun --

12:57:06   14          THE COURT:  The smoking gun.

12:57:08   15          MR. MOGIN:  A smoking gun.

12:57:08   16          THE COURT:  The smoking gun.

12:57:10   17  BY MR. MOGIN:

12:57:12   18  Q.  -- (continuing) one would have to anticipate the phrase in

12:57:14   19  advance in a Boolean search methodology?

12:57:18   20  A.  If you're using a Boolean search, you would either have to

12:57:22   21  anticipate the phrase, or I guess you could anticipate some

12:57:28   22  characteristic in metadata too.  The point being that the --

12:57:32   23  it's the anticipation that's difficult given the wide range of

12:57:34   24  possibilities.  The advantage supervised learning has is being

12:57:38   25  able to see the pattern from examples.

12:57:42    1    Q.  How could you formulate a realistic Boolean query for

12:57:50    2    "they are with us"?

12:57:50    3    A.  If you knew exactly that phrase, they are with us, some

12:57:54    4    Boolean languages would just let you put it in quotation

12:57:56    5    marks, other ones you would do a proximity, this word within

12:58:02    6    this proximity to this word.

12:58:02    7            So, I mean, if you knew exactly that phrase in

12:58:04    8    advance, you could bring it back, but the point is you don't

12:58:08    9    know exactly that phrase in advance.

12:58:10    10   Q.  So if one was unaware of the particular phrase that

12:58:14    11   constituted a smoking gun, would it be feasible or possible or

12:58:20    12   even likely to find it in a Boolean query?

12:58:26    13   A.  Could you repeat that question?  Sorry.  I didn't hear the

12:58:30    14   beginning.

12:58:32    15   Q.  If one did not know the precise phrasing of a smoking

12:58:36    16   gun-type document in advance, what would be the likelihood of

12:58:42    17   locating it with a Boolean query as compared to supervised

12:58:46    18   learning?

12:58:46    19   A.  Right.  So if by "smoking gun," we in particular mean a

12:58:50    20   document that uses coded or indirect language, given the vast

12:58:58    21   range of possibilities for that coded or indirect language, it

12:59:02    22   would be very difficult to use a Boolean query to find that,

12:59:04    23   and you would have more hope of doing that with supervised

12:59:08    24   learning with its sort of pattern-finding abilities.  It's not

12:59:12    25   going to be easy in either case.

| | | |
|---|---|---|
| 12:59:14 | 1 | Q. Now, let's go back to files for just a moment. When you |
| 12:59:20 | 2 | formulate a proposal or assist in the formulation of a |
| 12:59:24 | 3 | proposal, was there an underlying assumption that the |
| 12:59:30 | 4 | reviewing attorneys would be acting in good faith to find |
| 12:59:32 | 5 | responsive documents? |
| 12:59:32 | 6 | A. I am always assuming people who are reviewing documents |
| 12:59:38 | 7 | are acting in good faith. |
| 12:59:40 | 8 | Q. Well, did you plan for the possibility that the attorneys |
| 12:59:42 | 9 | would be trying to gain the system or cheating the system by |
| 12:59:46 | 10 | biassing the seed set, the training set? |
| 12:59:50 | 11 | A. No. I mean, my concern was with inadvertent bias that's |
| 12:59:56 | 12 | caused by knowing what the system did. I wasn't -- I wasn't |
| 01:00:00 | 13 | asked to contemplate any possibility of bad faith. |
| 01:00:04 | 14 | Q. Okay. Now, would it be your understanding as an |
| 01:00:16 | 15 | e-discovery information retrieval expert that the attorneys |
| 01:00:20 | 16 | who were making the evaluation for the training set were |
| 01:00:24 | 17 | subject to some sort of ethical or legal requirements with |
| 01:00:30 | 18 | respect to their findings regarding responsiveness? |
| 01:00:36 | 19 | THE COURT: Wait. I need to hear the end of that |
| 01:00:36 | 20 | question. Carolyn? |
| 01:00:56 | 21 | (Record read.) |
| 01:00:58 | 22 | THE COURT: Regarding the software? |
| 01:00:58 | 23 | MR. MOGIN: Regarding responsiveness. |
| 01:01:02 | 24 | THE COURT: Regarding responsiveness. |
| 01:01:04 | 25 | THE WITNESS: Well, I'm not a lawyer, but my |

01:01:06   1   understanding is that lawyers have certain ethical

01:01:10   2   obligations.  I don't know the legal codes in detail to

01:01:14   3   specify those.

01:01:16   4   BY MR. MOGIN:

01:01:36   5   Q.  Let's go back to recall if we can for just a moment.  In a

01:01:50   6   case of this nature, would there be some expectation about the

01:01:52   7   percentage of documents that would likely be recalled?  In

01:01:56   8   other words, what would be an acceptable, in your view, level

01:01:58   9   of recall?

01:02:02   10          MR. McKEOWN:  Objection.  Asked and answered.  He

01:02:04   11   says he has no opinion on this, your Honor.

01:02:06   12          THE COURT:  I don't think anybody asked him quite

01:02:08   13   that directly, so I am going to overrule your objection.

01:02:14   14          You can answer that, Doctor, if you know.

01:02:16   15          THE WITNESS:  I don't have an opinion of what an

01:02:20   16   appropriate level of recall is.  That seems to me to be a

01:02:26   17   legal question.

01:02:28   18   BY MR. MOGIN:

01:02:30   19   Q.  Well, in terms of percentages, do you think that 10

01:02:36   20   percent recall would be an acceptable number?

01:02:40   21          MR. McKEOWN:  Foundation, your Honor.  He just said

01:02:42   22   he doesn't know.

01:02:46   23          THE COURT:  Mr. Mogin, this has really got me

01:03:06   24   stumped.  First of all, Mr. Lewis just said, honestly, he

01:03:08   25   doesn't know.

01:03:10   1          MR. MOGIN:  I withdraw it.

01:03:12   2          THE COURT:  And we don't have another case that I

01:03:14   3   know of where the recall level was the issue in a word search.

01:03:18   4          MR. MOGIN:  I will withdraw the question.

01:03:20   5          THE COURT:  So I don't know what we would be

01:03:22   6   comparing it to.  I mean, that's one of the problems here.

01:03:32   7          MR. MOGIN:  Question withdrawn.

01:03:36   8          THE COURT:  Thank you.

01:03:36   9   BY MR. MOGIN:

01:03:36  10   Q.  Dr. Lewis, would you please look at Plaintiffs'

01:03:40  11   Exhibit 10.  Or, rather, yes, Plaintiffs' Exhibit 10.  It's

01:03:44  12   the proposed protocol.

01:03:44  13   A.  Yes.

01:03:46  14   Q.  And if you would please go to step 6.

01:03:58  15   A.  Okay.

01:04:00  16   Q.  Now, you recall that Mr. Neuwirth badgered you with a

01:04:02  17   number of questions --

01:04:02  18          THE COURT:  All right.  That's --

01:04:04  19   BY MR. MOGIN:

01:04:06  20   Q.  -- regarding --

01:04:08  21          THE COURT:  He vigorously cross-examined you,

01:04:10  22   correct.  Okay.

01:04:12  23   BY MR. MOGIN:

01:04:14  24   Q.  Well --

01:04:14  25          THE COURT:  I mean, it's all a matter of what you

01:04:16  1   call it, it is.

01:04:18  2         MR. MOGIN:  Without intending to insult the state

01:04:20  3   animal of Wisconsin.

01:04:24  4   BY MR. MOGIN:

01:04:24  5   Q.  In any event, you recall the questions that you were asked

01:04:28  6   that basically were to the effect that if the reviewing

01:04:32  7   attorneys game the system in some way, then the system

01:04:36  8   wouldn't be able to respond properly.  Do you recall that line

01:04:38  9   of questioning?

01:04:38  10  A.  I don't believe he used the word "game" --

01:04:42  11  Q.  He did not.

01:04:42  12  A.  -- but there was --

01:04:44  13        THE COURT:  Then why don't you use closer to what he

01:04:46  14  said.  Okay?

01:04:48  15  BY MR. MOGIN:

01:04:48  16  Q.  If the review was conducted in a biased fashion so that

01:04:50  17  because of responsiveness decisions, if that were to occur,

01:05:02  18  the point that he was trying to make was that no system would

01:05:04  19  be able to essentially recover from that, to be fully

01:05:10  20  utilized.  In other words, it would still depend on the front

01:05:12  21  end on good-faith decisionmaking in reviewing the terms?

01:05:20  22  A.  I wasn't entirely clear what point he was making, but my

01:05:24  23  testimony was that the system will attempt, to the extent it

01:05:30  24  can, to reproduce the pattern of responsiveness assessments

01:05:36  25  made by the people training it.

01:05:36 1 Q. Okay. And go to step 6, if you would, please, of

01:05:40 2 Exhibit 10.

01:05:42 3 A. Um-hmm.

01:05:42 4 Q. And one of the sentences that you were not asked to read

01:05:48 5 is that, Some or all of the non-privileged documents you added

01:05:54 6 to the training set according to the best practices, correct?

01:05:58 7 A. Yes.

01:05:58 8 Q. As well as sharing them with the plaintiffs. Did I read

01:06:02 9 that part of it correctly?

01:06:04 10 A. Yes, you did.

01:06:04 11 Q. Okay. And why did you include the sharing with the

01:06:08 12 plaintiffs in step 6?

01:06:10 13 A. Well, my understanding of the legal context is that both

01:06:18 14 defendants and plaintiffs have notions of responsiveness, that

01:06:24 15 there might be some disagreement between those two notions of

01:06:28 16 responsiveness, and that it would be important to have an

01:06:34 17 attempt to -- to the extent possible to receive it, to achieve

01:06:38 18 a consistent notion of responsiveness between the two so that

01:06:42 19 the statistical measures of, say, a recall would mean the same

01:06:48 20 thing to both sides.

01:06:50 21 Q. So --

01:06:52 22        MR. McKEOWN: Your Honor, I move to strike. It's

01:06:54 23 beyond the scope. Foundation. He is testifying as to legal

01:06:58 24 principles.

01:07:00 25        MR. MOGIN: Your Honor, the question was about what's

01:07:02    1   in step 6 of the protocol and why it was put in there.

01:07:04    2         THE COURT: Well, I am going to overrule your

01:07:06    3   objection. I think that is. And that was a -- that was part

01:07:10    4   of the protocol that wasn't asked about before.

01:07:16    5         We don't know whether Mr. Mogin put it in there or

01:07:20    6   did Dr. Lewis put it in there, but Dr. Lewis has sort of

01:07:24    7   adopted, I think he's kind of adopted -- he at least is aware

01:07:30    8   of what's in the protocol anyway.

01:07:30    9         So Mr. Mogin can ask that. We are not going to put

01:07:34   10   Mr. Mogin on the stand to say what he believed the protocol

01:07:40   11   was, so we will do it with Dr. Lewis.

01:07:42   12         MR. MOGIN: Could I have the last answer read back,

01:07:44   13   please?

01:08:18   14    (Record read.)

01:08:18   15         THE COURT: That's a long question. Okay.

01:08:20   16         MR. MOGIN: Actually, that was the answer.

01:08:22   17         THE COURT: Okay. Mr. Mogin, you helped me out. We

01:08:28   18   got through a half hour sort of like pre-lunch here. You are

01:08:34   19   going to get to do this again. Mr. Regard I hate to come back

01:08:38   20   a second time, I want to make sure we get Mr. Regard on.

01:08:44   21         MR. MOGIN: So you want me to stop?

01:08:46   22         THE COURT: Yes, if you don't mind. Okay?

01:08:46   23         And if Dr. Lewis -- I am saying that if Dr. Lewis has

01:08:50   24   any comments he wants to make of Mr. Regard's testimony,

01:08:54   25   hopefully, we're going to be able to do this in plenty of time

01:08:58   1   so that we can stop at 10 to 3:00.

01:09:04   2             MR. NEUWIRTH:  Your Honor --

01:09:04   3             MR. MOGIN:  May I ask two short questions before we

01:09:06   4   do a break, please, your Honor?

01:09:08   5             THE COURT:  Sure.

01:09:08   6   BY MR. MOGIN:

01:09:08   7   Q.  Dr. Lewis, you made a statement during direct that in your

01:09:14   8   opinion, it would take about, in some systems, one minute to

01:09:20   9   impact a million documents?  Do you recall that?

01:09:22   10  A.  Yes.

01:09:22   11  Q.  And there was some frivolity in the courtroom.  Could you

01:09:26   12  explain that, please?

01:09:26   13  A.  Well, you know, it obviously depends on things like the

01:09:30   14  length of the documents and the amount of computing power you

01:09:36   15  have available.  But I am certainly aware of figures like that

01:09:40   16  for the large search engine companies.  You know, what the

01:09:46   17  exact throughput is for e-discovery platforms, I am really not

01:09:52   18  sure, and that would depend on the power of the computers and

01:09:56   19  the size of the memory and the particular software and

01:09:58   20  whatnot.

01:10:00   21        So, yeah, it's a -- you know, one would have to

01:10:04   22  determine for an actual e-discovery system how fast they can

01:10:08   23  index it.

01:10:08   24  Q.  But it's your understanding that the commercial search

01:10:14   25  engines like Google, Yahoo, et cetera, are capable of doing

01:10:18  1  that?

01:10:18  2  A.  Yes, that's correct.

01:10:18  3  Q.  My last question is this.  Isn't it true that overall, the

01:10:26  4  key point in the validation process is recall?

01:10:30  5  A.  I would say key effectiveness measure in e-discovery is

01:10:36  6  recall because of the need to find a substantial proportion of

01:10:40  7  the responsive documents.

01:10:40  8  Q.  Okay.  So what's the confidence level on the recall of any

01:10:44  9  of the defendants in this case?

01:10:48  10  A.  There is no way you can tell.

01:10:48  11          MR. MOGIN:  Thank you.

01:10:50  12          THE COURT:  Mr. Neuwirth.

01:10:52  13          MR. NEUWIRTH:  May I just cover one item that was

01:10:54  14  directly raised in Mr. Mogin's examination?

01:10:58  15                          - - -

01:10:58  16          DANIEL D. LEWIS, RECROSS-EXAMINATION

01:10:58  17  BY MR. NEUWIRTH:

01:10:58  18  Q.  Dr. Lewis, you will recall that Mr. Mogin focused in his

01:11:04  19  questioning at the beginning on my March 21st letter which we

01:11:08  20  had gone over today that I believe we marked as Exhibit 9, and

01:11:16  21  Mr. Mogin --

01:11:16  22  A.  Excuse me.  I have an Exhibit 10, which is a March 21st

01:11:22  23  letter.

01:11:22  24  Q.  Exhibit 10, thank you.

01:11:24  25          And Mr. Mogin asked you about the information we went

01:11:28   1   over, that there were 660 documents in the sample set, that 27

01:11:34   2   of those were found to be at least marginally responsive, that

01:11:38   3   this was done to have a 99 percent confidence level and a 5

01:11:42   4   percent margin of error.

01:11:44   5         And Mr. Mogin asked you, he said, You didn't have all

01:11:46   6   that information in the March 21st letter at the time you

01:11:50   7   first testified, right?

01:11:50   8   A.   Actually, I don't recall if he is asking about the last

01:11:56   9   set of 660, which had 28 responsive, or the first set that had

01:12:02   10  27 responsive, or the middle set that had 31.

01:12:06   11  Q.   But the idea -- but he asked you whether information about

01:12:12   12  that type of confidence level, that margin of error, the size

01:12:16   13  of the sample set, and the number of documents that had been

01:12:18   14  found, he asked you whether you had that information when you

01:12:20   15  last testified, and you said no?

01:12:22   16  A.   Yeah, that's right, I did not have this information about

01:12:26   17  the size of the hits, the size of the null sets, and all of

01:12:30   18  the details of this last evaluation.

01:12:34   19  Q.   I would just ask you to take out Plaintiffs' Exhibit 4,

01:12:38   20  which is the November 22nd, 2011, letter that you were also

01:12:42   21  asked about that I sent to plaintiffs.

01:12:50   22         THE COURT:  Which is --

01:12:50   23         MR. NEUWIRTH:  It's Plaintiffs' Exhibit 4.

01:12:52   24         THE WITNESS:  So I have Plaintiffs' 4, yes.

01:12:54   25  BY MR. NEUWIRTH:

01:12:54　　1　Q.　Okay.　And I have a copy --

01:12:56　　2　　　　　　MR. NEUWIRTH:　I have an extra copy for your Honor.

01:12:58　　3　　　　　　THE COURT:　I have it here.

01:13:00　　4　BY MR. NEUWIRTH:

01:13:00　　5　Q.　I'd like you to turn to the bottom of page 3.

01:13:02　　6　A.　Okay.

01:13:02　　7　Q.　This is from November 22nd, many months before you

01:13:06　　8　testified.　Look at the last bullet on page 3.

01:13:12　　9　Georgia-Pacific, in consultation with Counsel on Call, sought

01:13:16　10　to validate the null set to a 99 percent confidence level and

01:13:22　11　a 5 percent margin of error by reviewing and vetting of 660

01:13:26　12　randomly selected documents, the set validated with 27

01:13:32　13　marginally responsive documents.

01:13:34　14　　　　　So that information was in a November 22nd letter.

01:13:38　15　It was not provided for the first time on March 21st, correct?

01:13:42　16　A.　The size of the null set is not here, the size of the hit

01:13:48　17　set is not here, the actual statistic which the confidence

01:13:56　18　interval is being computed for is not here.

01:13:58　19　Q.　Right.　So the results are here, and there was

01:14:00　20　clarification in the March 21st letter, correct?

01:14:04　21　A.　That's an interesting interpretation.

01:14:08　22　　　　　　MR. NEUWIRTH:　Thank you.

01:14:12　23　　　　　　THE COURT:　Okay.　Time to eat.　Can you be back here

01:14:16　24　at a quarter to 1:00?

01:14:18　25　　　　　Yes.

01:14:18   1       MR. ECHOLS: One totally housekeeping matter, Judge.

01:14:22   2  Barack Echols. We had marked Defendants' Exhibit 9 as the

01:14:26   3  Blair and Maron study, and then that affidavit of Mr. Clancy

01:14:30   4  was tendered, and it hadn't actually been marked. I didn't

01:14:32   5  know if you wanted to have a number on it.

01:14:34   6       THE COURT: Are you going to move -- yes, well, let's

01:14:36   7  mark it anyway.

01:14:38   8       MR. ECHOLS: And 10 was this March --

01:14:40   9       THE COURT: Are you keeping track, I hope, because I

01:14:42  10  am not?

01:14:42  11       MR. ECHOLS: Would you prefer that that be 9-A or be

01:14:46  12  marked as Exhibit 11?

01:14:48  13       MR. MAROVITZ: I premarked.

01:14:54  14       THE COURT: Do you have a preference?

01:14:54  15       MR. MOGIN: None at all.

01:14:56  16       THE COURT: I am just glad somebody is doing it.

01:14:58  17  Thank you.

01:15:02  18       MR. ECHOLS: We will make it 9-A.

01:15:04  19       THE COURT: Let's make it a quarter to 1:00 as

01:15:06  20  possible. That will give us two hours then.

01:15:08  21       MR. MOGIN: We are done with Dr. Lewis.

01:15:10  22       THE COURT: Dr. Lewis, we are finished with you for

01:15:12  23  now. I don't know whether you will want to come back or not.

01:15:16  24  We are starting with Mr. Regard.
          (The trial was adjourned at 12:15 p.m. until 12:45 p.m. of
 25  this same day and date.)

```
1                    IN THE UNITED STATES DISTRICT COURT
                        NORTHERN DISTRICT OF ILLINOIS
2                            EASTERN DIVISION

3
     KLEEN PRODUCTS, LLC, et al.,        )  Docket No. 10 C 5711
4                                        )
                            Plaintiffs,  )
5                                        )
                    vs.                  )
6                                        )
     PACKAGING CORPORATION OF AMERICA,   )  Chicago, Illinois
7    et al.,                             )  March 28, 2012
                                         )  1:45 o'clock p.m.
8                           Defendants.  )

9
                TRANSCRIPT OF PROCEEDINGS - EVIDENTIARY HEARING
10          BEFORE THE HONORABLE MAGISTRATE JUDGE NAN R. NOLAN
                             VOLUME 2-B
11

12   APPEARANCES:

13
     For the Plaintiffs:        THE MOGIN LAW FIRM
14                              BY:  MR. DANIEL J. MOGIN
                                707 Broadway, Suite 1000
15                              San Diego, CA  92101
                                (619) 687-6611
16
                                FREED KANNER LONDON & MILLEN LLC
17                              BY:  MR. MICHAEL J. FREED
                                MR. ROBERT J. WOZNIAK
18                              2201 Waukegan Road, Suite 130
                                Bannockburn, IL  60015
19                              (224) 632-4500

20                              SCOTT & SCOTT
                                BY:  MR. WALTER W. NOSS
21                              707 Broadway, Suite 1000
                                San Diego, CA  92101
22                              (619) 233-4565

23
     Court Reporter:           MS. CAROLYN R. COX, CSR, RPR, CRR, FCRR
24                              Official Court Reporter
                                219 S. Dearborn Street, Suite 1854-B
25                              Chicago, Illinois  60604
                                (312) 435-5639
```

```
 1   APPEARANCES CONTINUED:

 2   For the Plaintiffs:        LOCKRIDGE GRINDAL NAUEN PLLP
                                BY:  MR. BRIAN D. CLARK
 3                              100 Washington Avenue S.
                                Minneapolis, MN  55401
 4                              (612) 239-6900

 5
     For Defendant Packaging    KIRKLAND & ELLIS LLP
 6   Corporation of America:    BY:  MR. BARACK S. ECHOLS
                                     MR. LEONID FELLER
 7                              300 North LaSalle Street
                                Chicago, IL  60654
 8                              (312) 862-3144

 9
     For Defendant             FOLEY & LARDNER LLP
10   International Paper:       BY:  MR. JAMES T. McKEOWN
                                     MR. NATHAN EIMER
11                              777 East Wisconsin Avenue
                                Milwaukee, WI  53202
12                              (414) 297-5530

13
                                FOLEY & LARDNER LLP
14                              BY:  MS. JOANNE LEE
                                321 North Clark Street, Suite 2800
15                              Chicago, IL  53202

16   For Defendant             MAYER BROWN LLP
     Temple-Inland:            BY:  MR. ANDREW S. MAROVITZ
17                                   MS. BRITT M. MILLER
                                71 South Wacker Drive
18                              Chicago, IL  60606
                                (312) 782-0600

19
     For Defendant             K & L GATES LLP
20   Cascades and Norampac:     BY:  MR. SCOTT M. MENDEL
                                70 West Madison Street, Suite 3100
21                              Chicago, IL  60602
                                (312) 372-1121

22

23   For Defendant             QUINN EMANUEL URQUHART &
     Georgia-Pacific:          SULLIVAN LLP
24                              BY:  MR. STEPHEN R. NEUWIRTH
                                51 Madison Avenue, 22nd Floor
25                              New York, NY  10010
                                212-849-7000
```

```
 1   APPEARANCES CONTINUED:

 2   For Defendant              WINSTON & STRAWN LLP
     RockTenn CP, LLC:          BY:  MR. R. MARK McCARIENS
 3                                   MR. JOSEPH L. SIDERS
                                35 West Wacker Drive
 4                              Chicago, IL  60601
                                (312) 558-5902
 5

 6   For Defendant              McDERMOTT WILL & EMERY LLP
     Wayerhaeuser Company:      BY:  MS. RACHAEL V. LEWIS
 7                              600 13th Street NW
                                Washington, DC  2005
 8                              (202) 756-8479

 9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25
```

1    (The following proceedings were had in open court:)

01:53:52    2        THE COURT:  Are you ready to call Mr. Regard?

01:53:54    3        MR. MAROVITZ:  I am, your Honor.  Thank you.

01:53:54    4    (Witness sworn.)

01:54:18    5        MR. MAROVITZ:  Your Honor, before I start, the first

01:54:20    6    exhibit that we are going to use is Defense Exhibit 11, which

01:54:22    7    is the report of Daniel Regard.  We didn't bring extra copies,

01:54:28    8    we sent so many around.

01:54:30    9        MR. MOGIN:  Is the report coming into evidence at

01:54:30    10   this point, your Honor?

01:54:32    11       MR. MAROVITZ:  I didn't ask that it be admitted.

01:54:36    12   It's just for identification purposes.

01:54:38    13       THE COURT:  He may need it to refresh recollection or

01:54:40    14   do whatever.  That's fine.  We all have -- everyone in the

01:54:42    15   courtroom who is part of the case I assume has a copy.  Thank

01:54:46    16   you.

01:54:48    17       MR. MAROVITZ:  Your Honor, do you mind if I approach?

01:54:50    18       THE COURT:  Of course.

01:54:50    19       MR. MAROVITZ:  Thank you.  Thank you, your Honor.

01:55:00    20                         - - -

01:55:00    21       DAN REGARD, DIRECT EXAMINATION CONTINUED

01:55:00    22   BY MR. MAROVITZ:

01:55:02    23   Q.  Welcome back.  State your name for the record.

01:55:04    24   A.  My name is Dan Regard.

01:55:06    25   Q.  With the Court's indulgence, I'd like to set the table

| | | |
|---|---|---|
| 01:55:10 | 1 | here very quickly to provide some context for your testimony |
| 01:55:14 | 2 | today, Mr. Regard, and to avoid undue repetition from your |
| 01:55:16 | 3 | testimony on February 21st. |
| 01:55:18 | 4 | So back on February 21st when you began your |
| 01:55:22 | 5 | testimony, you described your opinions, your qualifications, |
| 01:55:28 | 6 | your education, and your employment, and the work that you did |
| 01:55:30 | 7 | to familiarize yourself with defendants' discovery protocols, |
| 01:55:34 | 8 | correct? |
| 01:55:34 | 9 | A. Yes, sir. |
| 01:55:34 | 10 | Q. And you explained in your view ESI best practices and told |
| 01:55:40 | 11 | us the key words were very well accepted in ESI circles and, |
| 01:55:44 | 12 | when applied properly, could lead to a robust process and the |
| 01:55:48 | 13 | high quality results, correct? |
| 01:55:50 | 14 | A. That is correct. |
| 01:55:50 | 15 | Q. And to follow up on all of that, you submitted an expert |
| 01:55:54 | 16 | report on Wednesday, March 21st, correct? |
| 01:55:58 | 17 | A. Yes, sir. |
| 01:56:00 | 18 | Q. And I've handed you what's been marked as Defense |
| 01:56:08 | 19 | Exhibit 11 for identification a copy of that expert report. |
| 01:56:12 | 20 | Is that a true and accurate copy of your expert report in this |
| 01:56:14 | 21 | matter? |
| 01:56:14 | 22 | A. Yes, it is. |
| 01:56:18 | 23 | MR. MAROVITZ: Judge, just as an aside, I am sure we |
| 01:56:22 | 24 | can scare one up if you need one. |
| 01:56:22 | 25 | THE COURT: I have one. Thank you. |

01:56:24  1  BY MR. MAROVITZ:

01:56:24  2  Q.  Now, that expert report at 1 describes the four opinions

01:56:28  3  that you issued during your February testimony, correct?

01:56:32  4  A.  Yes, sir.

01:56:32  5  Q.  Okay.  I'd like to start drilling down on the first

01:56:38  6  opinion that you already offered back in February by focusing

01:56:40  7  on the use of key words as part of the search methodology.

01:56:44  8        What, in your view, are the best practices for using

01:56:48  9  key words as part of your review process?

01:56:50  10  A.  Well, it's --

01:56:52  11        MR. MOGIN:  Your Honor, may I just remind the court

01:56:54  12  that we have a continuing objection with regard to best

01:56:56  13  practices testimony here?

01:56:58  14        THE COURT:  Yes.  Thank you.

01:57:00  15        Well, do you want to say -- do you want to state

01:57:04  16  again what your objection is based on best practices.

01:57:06  17        MR. MOGIN:  Yes, your Honor.  Particularly -- this is

01:57:08  18  particularly important with respect to Mr. Regard, who has

01:57:12  19  testified that he is an attorney by training and that the

01:57:16  20  issue is having best practices testimony come in or vice

01:57:24  21  versa.  It's having legal opinions come in disguised as best

01:57:30  22  practices testimony.  It's precisely the matter that's

01:57:32  23  described in the first motion we filed before the first

01:57:36  24  hearing, and you -- if you recall, your Honor, you have had

01:57:40  25  some reservations, but there are certain aspects of the

01:57:44　1　testimony that Mr. Regard got into talking about best

01:57:48　2　practices, you had some reservations about whether we should

01:57:50　3　be going in that area or not.

01:57:54　4　　　　　THE COURT:  I don't recall that; but yesterday, when

01:57:58　5　I got your motion and I was sitting thinking about it, I think

01:58:06　6　Mr. Regard is here as a lawyer, but as far as I know, most of

01:58:14　7　his business is he is an e-discovery consultant and happens to

01:58:22　8　be a lawyer.

01:58:28　9　　　　　So I know that you have multiple objections to

01:58:32　10　Mr. Regard, just like Dr. Lewis is very definitely quite

01:58:38　11　knowledgeable in statistics, in search retrieval, and he's

01:58:44　12　also been asked in this specific case to do -- to put on a

01:58:50　13　third hat, if you will, or at least a second hat.

01:58:54　14　　　　　So, Mr. Regard, as far as I am sitting here, you

01:58:58　15　know, I think he is also in the bucket of e-discovery expert.

01:59:06　16　　　　　Now, if you don't, then I guess you will talk to him

01:59:10　17　about that.

01:59:10　18　　　　　MR. MOGIN:  I certainly will, your Honor, but the

01:59:12　19　point has to do with best practices --

01:59:16　20　　　　　THE COURT:  Don't you think people -- consultants do

01:59:20　21　know best practices, whether they're lawyers or not?  I mean,

01:59:24　22　you're going to have to talk to me more specifically.

01:59:26　23　　　　　MR. MOGIN:  Well, we are going to get right back to

01:59:28　24　the example that I gave you before, your Honor.  If we were in

01:59:32　25　a criminal proceeding, Miranda would be a statement of the

01:59:38  1  law.

01:59:38  2  THE COURT:  Right.

01:59:38  3  MR. MOGIN:  How the police effectuate Miranda, did

01:59:42  4  you give a Miranda warning, might be best practices or come

01:59:46  5  within some testimony by an expert on police procedures

01:59:50  6  regarding best practices.  But the fact of the need to give

01:59:54  7  the Miranda warning is a matter of law.

01:59:58  8  Now, if we are going to have someone, particularly

02:00:00  9  someone who is a lawyer, give best practices testimony about

02:00:06  10  best practices in discovery when it's you, your Honor, who

02:00:10  11  ultimately makes a determination about whether defendants have

02:00:14  12  complied with or have done a reasonable search here, we can't

02:00:20  13  -- it's difficult if the lines get crossed.

02:00:28  14  THE COURT:  I think Dr. Lewis for two days said that

02:00:32  15  they didn't.

02:00:32  16  MR. MAROVITZ:  He also said --

02:00:34  17  THE COURT:  I'm sorry.  But, you know, I don't know

02:00:38  18  whether he used those exact words, but it sure was the input.

02:00:44  19  MR. MOGIN:  Well --

02:00:44  20  THE COURT:  I mean, I can't do this outside of the

02:00:48  21  context of a specific question, so I think what we ought to do

02:00:52  22  is get to a specific question; I am inviting you to jump up on

02:00:58  23  a specific question and let's deal with it.  Okay?

02:01:00  24  MR. MOGIN:  Very good.

02:01:00  25  THE COURT:  All right.

02:01:02  1  BY MR. MAROVITZ:

02:01:04  2  Q.  Mr. Regard, what are your best practices for using key

02:01:14  3  words other than due processing?

02:01:14  4  A.  In my experience, the best practices for using key words

02:01:16  5  or really any technology are to have a thoughtful process in

02:01:22  6  the front end where there's substantial thought and research

02:01:24  7  that goes into developing your approach; it's to have an

02:01:28  8  iterative process to revise and enhance, improve your

02:01:34  9  application; it's to communicate with opposing parties or

02:01:40  10  counter-parties the process you're going through; it's to

02:01:42  11  apply your process, your key words, in this case, to the

02:01:46  12  actual corpus of documents and to evaluate the results.

02:01:50  13  Q.  And by evaluating the results, what do you mean by that?

02:01:56  14  A.  Well, in this case, it was the -- after key words were

02:02:04  15  applied, there was an evaluation done of the application of

02:02:08  16  those key words in particular by testing the documents that

02:02:10  17  were left behind that we have referred to so often as the null

02:02:16  18  set to see if the results in the null set or the results that

02:02:20  19  you would find in null sets are consistent with your

02:02:22  20  expectations that you had done a good job.

02:02:24  21  Q.  For purposes of reaching your opinion in this case that

02:02:28  22  the search methodology undertaken by the defendants meets or

02:02:32  23  exceeds best practices, I'd like to walk through what you

02:02:38  24  relied upon regarding what the defendants did.  Okay?

02:02:40  25  A.  Okay.

02:02:42  1  Q.  So first, do you understand that defendants, to use the

02:02:48  2  first element you talked about a minute ago, implemented a

02:02:52  3  well thought-out process with substantial input from the

02:02:54  4  lawyers and the business people?

02:02:56  5  A.  That's my understanding from my review of the materials

02:02:58  6  and my discussions, yes.

02:03:02  7  Q.  And what exactly did you review, and how did you come to

02:03:04  8  that understanding?

02:03:04  9  A.  What I reviewed was the descriptions in correspondence, in

02:03:14  10  pleadings, of the process the defendants went through in

02:03:16  11  developing their key words from the beginning; added to that

02:03:22  12  were my discussions with individual defendants and their

02:03:24  13  counsel and the consultants they work with; to understand the

02:03:34  14  process they went through at the very beginning of 2011 when

02:03:36  15  they received the request for the production from the

02:03:38  16  plaintiffs on how to develop an approach -- and in this case,

02:03:40  17  key words -- to respond to the document request they had been

02:03:44  18  delivered.

02:03:44  19  Q.  What next?

02:03:46  20  A.  Well, that was really the formation of the response.  And

02:03:52  21  there was a lot of testimony about what Georgia-Pacific ended

02:03:56  22  up doing.  That testimony was in February when we were last

02:04:00  23  here in court, and I can repeat some of that, but I don't know

02:04:02  24  that it's necessary in the interest of time.

02:04:06  25          Would you like me to go into more detail?

02:04:10  1  Q.  No, I don't think so.

02:04:10  2  A.  Okay.

02:04:10  3  Q.  So you had an understanding based upon the correspondence

02:04:14  4  and documents and the testimony of Mr. Koch and Mr. Brown and

02:04:20  5  Georgia-Pacific.  What happened next after the Georgia-Pacific

02:04:24  6  parties testified in court?

02:04:26  7  A.  Well, as we go through this process, I want to point one

02:04:28  8  thing out.  It talked about five different aspects of really

02:04:32  9  developing a robust process.  They don't necessarily happen in

02:04:36  10  sequence one after the other, so these can happen

02:04:38  11  simultaneously.  I just want to make sure that all five

02:04:42  12  elements are there:  that there was a thoughtful process,

02:04:44  13  there was iterative testing, there was communication, there

02:04:48  14  was application against real documents, and then there was an

02:04:50  15  evaluation at the end.

02:04:52  16          THE COURT:  All right.  I have a question right here.

02:04:54  17          THE WITNESS:  Yes, ma'am.

02:04:54  18          THE COURT:  I think you first said they developed a

02:05:00  19  word search, Georgia-Pacific developed a word search.  We

02:05:04  20  don't know -- while we sort of know from Brown and Koch how

02:05:08  21  they did it, we don't know anything how the other seven.  Did

02:05:12  22  they use the same words as Georgia-Pacific, because we have

02:05:18  23  heard you're it for the other six or seven?  So what do you

02:05:22  24  know about how they developed their word search?

02:05:26  25          THE WITNESS:  That's a great question, your Honor,

02:05:28  1  and I'm going to respond to that right now.

02:05:30  2         THE COURT:  Okay.

02:05:32  3         THE WITNESS:  Again, my understanding from what I

02:05:36  4  have read and what the discussion I have had with the

02:05:38  5  defendants is, they all participated in the development of the

02:05:42  6  initial search terms.  All seven defendants participated, even

02:05:46  7  though it was Georgia-Pacific that tested it against their

02:05:50  8  specific documents.  So they all had a chance to cooperate,

02:05:54  9  they all had a chance to give meaningful input, they all had a

02:05:58  10  chance to discuss with their own clients and the witnesses and

02:06:02  11  the employees from their clients as to what these requests

02:06:06  12  meant, how those requests might fit their organization, and

02:06:08  13  help create the initial set of search terms.

02:06:12  14         THE COURT:  So each of the defendants had the request

02:06:14  15  to produce before they started developing word search groups?

02:06:20  16         THE WITNESS:  They all received it at the same time

02:06:22  17  is my understanding, your Honor.

02:06:22  18         THE COURT:  Okay.  All right.

02:06:24  19         THE WITNESS:  And Georgia-Pacific took it upon

02:06:28  20  themselves on behalf of the entire group to first test these

02:06:32  21  search terms against their set of documents first, which ended

02:06:36  22  up -- and I think the earlier testimony was such -- that they

02:06:40  23  spent a lot of time and a lot of iterations and hundreds of

02:06:42  24  hours reviewing documents, testing search terms, revising the

02:06:46  25  search terms, until they got to a point where they felt they

| | | |
|---|---|---|
| 02:06:52 | 1 | were getting the type of performance that was acceptable given |
| 02:06:58 | 2 | the amount of effort that they were putting into the process. |
| 02:07:00 | 3 | THE COURT: Okay. |
| 02:07:00 | 4 | THE WITNESS: Then they tested that against a random |
| 02:07:04 | 5 | sample, Georgia-Pacific did, and came back with an answer of a |
| 02:07:08 | 6 | low residual number of responsive documents left behind in a |
| 02:07:14 | 7 | null set. And my understanding is that at that point, those |
| 02:07:18 | 8 | search terms were communicated to the plaintiffs. |
| 02:07:24 | 9 | THE COURT: Do you know when that was? |
| 02:07:24 | 10 | THE WITNESS: I believe it was in November, your |
| 02:07:28 | 11 | Honor, 2011, but I could have the date or the month wrong. |
| 02:07:34 | 12 | THE COURT: Okay. |
| 02:07:34 | 13 | BY MR. MAROVITZ: |
| 02:07:36 | 14 | Q. Can I suggest, Mr. Regard, it might have been August of |
| 02:07:38 | 15 | 2011? |
| 02:07:38 | 16 | A. It may have been August of 2011. That may be correct. |
| 02:07:42 | 17 | THE COURT: Okay. |
| 02:07:44 | 18 | THE WITNESS: There was a lot of activity back and |
| 02:07:46 | 19 | forth in the correspondence. |
| 02:07:48 | 20 | THE COURT: Okay. |
| 02:07:48 | 21 | THE WITNESS: At that point, a response was received |
| 02:07:50 | 22 | from the plaintiffs that said, These are the concerns we have |
| 02:07:54 | 23 | with your individual search terms for the individual requests, |
| 02:07:58 | 24 | and that the defendants, again led specifically by |
| 02:08:04 | 25 | Georgia-Pacific, but the defendants collectively revised those |

02:08:08 1  search terms, developed a second set, if you will. It's

02:08:12 2  really not the second set because in the first round, they had

02:08:18 3  devised many iterative sets, but the second major set of

02:08:22 4  search terms, and those were, again, tested against the

02:08:24 5  Georgia-Pacific documents. And when the revisions were

02:08:28 6  complete, the null set was created and the null set was tested

02:08:32 7  a second time.

02:08:36 8       Now, I want to get to your question, your Honor, what

02:08:38 9  did the individual defendants do? It was at this point that

02:08:42 10  the search terms were distributed to the individual defendants

02:08:46 11  for them to use on their own documents, and it was at this

02:08:50 12  point that the individual defendants modified the search

02:08:54 13  terms, and they modified them in a number of ways.

02:09:00 14       One way that they modified them was to remove

02:09:04 15  references to Georgia-Pacific-specific facilities and to

02:09:10 16  replace that with specific facilities or terms that were

02:09:14 17  specific to their organizations.

02:09:18 18       Another way that they changed those search terms was,

02:09:22 19  for example, when Georgia-Pacific was looking for

02:09:24 20  communications as they were requested to produce with the

02:09:28 21  codefendants, the other defendants would remove their name

02:09:34 22  from that list and replace it with Georgia-Pacific because

02:09:36 23  they weren't looking for communications with themselves.

02:09:40 24       And then a third way that they changed those search

02:09:42 25  terms was to modify them as appropriate for the technology

| | | |
|---|---|---|
| 02:09:46 | 1 | that they were using to manage and search their own documents. |
| 02:09:50 | 2 | BY MR. MAROVITZ: |
| 02:09:52 | 3 | Q.  Once those search strings were modified, was it one search |
| 02:09:54 | 4 | string per defendant? |
| 02:09:56 | 5 | A.  No, sir.  There were -- my recollection is about 15 search |
| 02:10:04 | 6 | strings in the original set responding to 15 different |
| 02:10:10 | 7 | document requests, but because of technology and complexity, |
| 02:10:14 | 8 | those ended up getting broken down into multiple strings.  In |
| 02:10:20 | 9 | the case of one defendant, it was as many as 41 different |
| 02:10:22 | 10 | search strings. |
| 02:10:24 | 11 | And within those search strings -- it wasn't just a |
| 02:10:28 | 12 | single phrase, your Honor.  So there was some testimony |
| 02:10:30 | 13 | earlier today that you could run a single search, like put in |
| 02:10:34 | 14 | a phrase in quotes.  Rather, these search strings contained |
| 02:10:40 | 15 | multiple terms and Boolean operators, Boolean ands and ors and |
| 02:10:48 | 16 | parentheses that actually had the result of creating many, |
| 02:10:54 | 17 | many, many more search terms.  Some of the search terms I |
| 02:10:58 | 18 | looked at had dozens of individual words, but because of the |
| 02:11:02 | 19 | way the Boolean operators worked, it would result in hundreds |
| 02:11:04 | 20 | of variations of searches being executed simultaneously. |
| 02:11:10 | 21 | And on top of that, compounded by that, different |
| 02:11:14 | 22 | defendants used technologies called either wild cards or |
| 02:11:18 | 23 | stemming that resulted in variations for each of the words in |
| 02:11:22 | 24 | the search strings.  Because of wild cards and stemmings, |
| 02:11:26 | 25 | search strings that had dozens of words and hundreds of |

02:11:28   1   permutations ended up having hundreds of thousands of

02:11:32   2   permutations.

02:11:34   3         When I look over the set of search strings, my

02:11:36   4   estimate is there were more than 2 million searches executed

02:11:40   5   simultaneously, which is a very complex set of searches.

02:11:44   6   Q. And you've made this judgment based upon your review of

02:11:46   7   the letters and your discussions with representatives of the

02:11:52   8   defendants as you described that, correct?

02:11:54   9   A. Well, I'm really not describing judgment at this point.

02:11:56   10   This is the information as I learned it. This is the sequence

02:12:00   11   of chronology.

02:12:00   12   Q. Did you have some personal involvement in some way in the

02:12:06   13   process as opposed to simply appearing here as an expert

02:12:08   14   witness?

02:12:10   15   A. Well, I was engaged in this case in 2010, very early on

02:12:18   16   when this case was first filed is my understanding, and I was

02:12:22   17   hired then by Temple-Inland to help them with collection

02:12:28   18   methodologies for files.

02:12:30   19         So my involvement in the collection process was very

02:12:32   20   early on, but my evaluation of the search process only began

02:12:38   21   in late 2011.

02:12:40   22   Q. And were you given free reign to ask questions of the

02:12:42   23   other defendants to make that evaluation?

02:12:46   24   A. I was given an opportunity to ask any questions I wanted

02:12:50   25   to, not once but twice, and I was able to speak with

02:12:52  1  representatives of both defense counsel and the consultants

02:12:58  2  for each one of the defendants.

02:12:58  3  Q.  One of the things you mentioned, Mr. Regard, was that set

02:13:02  4  of letters that you reviewed that enabled you to come to your

02:13:08  5  -- well, that enabled you, essentially, you reviewed in effort

02:13:12  6  to come to your opinions that you're expressing today.  We put

02:13:14  7  those together as a group exhibit, so I want to have these

02:13:20  8  marked, if I may, as --

02:13:24  9        THE COURT:  Is this the book you sent me?

02:13:26  10        MR. MAROVITZ:  Not quite.  Everything that we are

02:13:28  11  handing over was previously sent.  These were all the letters

02:13:30  12  that were dated March 21st, 2012.

02:13:32  13        THE COURT:  Okay.

02:13:34  14        MR. MAROVITZ:  We are marking them as Group Exhibit

02:13:36  15  12, and it's A through, I think, G.  In the event that you get

02:13:44  16  a set that's doesn't have A through G, they're alphabetical

02:13:48  17  with past status going first.  So I will hand that up.

02:13:48  18        MR. MOGIN:  Again, your Honor, I assume this is

02:13:50  19  marking strictly for identification because these are the same

02:13:54  20  letters that we have been talking about all morning long as

02:14:00  21  being hearsay.

02:14:02  22  BY MR. MAROVITZ:

02:14:12  23  Q.  Mr. Regard, are these the letters to which you're

02:14:14  24  referring that you were able to review before today?

02:14:16  25  A.  Well, I definitely did review these letters before today,

| | | |
|---|---|---|
| 02:14:22 | 1 | but the letters I was referring to in the earlier testimony |
| 02:14:26 | 2 | were the letters and correspondence between the parties during |
| 02:14:30 | 3 | 2011 and the early part of 2012.  And those were the letters |
| 02:14:36 | 4 | that described the process that was used to develop the search |
| 02:14:40 | 5 | terms, the testing, the disclosure of the search terms, and |
| 02:14:44 | 6 | the response from the plaintiffs to the initial set of search |
| 02:14:48 | 7 | terms. |
| 02:14:48 | 8 | Q.  Now, after each defendant customized its search strings, |
| 02:14:54 | 9 | as you mentioned, what happened next? |
| 02:14:56 | 10 | A.  Well, after they customized the search, while this process |
| 02:15:00 | 11 | of search string development was going on, the defendants were |
| 02:15:04 | 12 | also collecting documents; they were identifying key players, |
| 02:15:10 | 13 | they were collecting documents from those key players in other |
| 02:15:14 | 14 | locations, and they were preparing those documents.  And so |
| 02:15:16 | 15 | when the search terms were modified and customized for each |
| 02:15:22 | 16 | defendant, they then applied them against the documents they |
| 02:15:26 | 17 | had collected. |
| 02:15:26 | 18 | Q.  Didn't the defendants use the search strings to collect |
| 02:15:30 | 19 | their documents in the first place? |
| 02:15:32 | 20 | A.  No, they collected documents on a wholesale basis by |
| 02:15:34 | 21 | custodian.  They identified key players in the organization, |
| 02:15:40 | 22 | which was also part of the correspondence back and forth.  And |
| 02:15:44 | 23 | once they collected those documents, it was these wholesale |
| 02:15:48 | 24 | collections by custodian that they then applied the search |
| 02:15:52 | 25 | terms against. |

02:15:52 1 Q. And after they applied the search terms against those

02:15:56 2 collections by custodian, what came next?

02:16:00 3 A. Well, by definition, to be clear, when you apply the

02:16:06 4 search terms against a set of documents, you get a set of

02:16:10 5 hits. And for every set of hits, you have a corresponding

02:16:12 6 non-hit set, which we call the null set.

02:16:16 7 So when they applied the search terms that they had

02:16:20 8 modified to the documents they had collected, they then went

02:16:24 9 to the null set, drew forth a random sample of documents, and

02:16:28 10 reviewed them.

02:16:30 11 Then they -- shall I continue?

02:16:32 12 Q. Please.

02:16:34 13 A. Then they made -- they conducted a measurement of what is

02:16:38 14 the ratio of the responsive documents in the null set to the

02:16:42 15 number of documents they tested in the null set and used that

02:16:46 16 to create an estimate of the number of documents that might be

02:16:48 17 residual in the null set.

02:16:50 18 Q. If you're using the right methodology in the first place,

02:16:54 19 why did you expect to have any documents in the null set that

02:16:58 20 would be responsive?

02:16:58 21 A. That's just one of the realities of document management

02:17:04 22 and information retrieval is that one of the reasons we have

02:17:10 23 the science of information retrieval is to improve our methods

02:17:14 24 of retrieving documents against an aspiration to try to do it

02:17:18 25 as well as possible but a reality that it's not going to be

02:17:22  1  perfect.

02:17:22  2      The question is not, was it perfect.  At least in the

02:17:26  3  cases I have worked in over the last 15, 20 years, the

02:17:28  4  question is, how imperfect is it and what is the ratio, what

02:17:32  5  is the proportionality of trying to work harder to get a

02:17:36  6  better result versus the cost, time, and effort to do that.

02:17:38  7  Q.  We haven't heard a lot about proportionality in this

02:17:42  8  hearing so far.  What do you mean by that?

02:17:42  9  A.  Well, every time you look at -- in this case, you look at

02:17:48  10  a set of search terms, you apply it against the corpus of

02:17:52  11  documents, you get results, you evaluate the results.  A

02:17:56  12  decision has to be made, can we look at more sample documents,

02:18:00  13  can we learn more about the documents, can we revise the

02:18:02  14  search terms, can we run the tests again, how long does that

02:18:06  15  take, how much money does that cost, and what are the

02:18:10  16  improvements we are getting in the system for exercising that

02:18:12  17  time and expense and effort.

02:18:14  18      Eventually, there's a proportionality of do we stop

02:18:18  19  trying to make it marginally better and spending more money,

02:18:24  20  or do we start reviewing documents and producing them to the

02:18:26  21  opposing party.

02:18:28  22      MR. MOGIN:  Popping up, your Honor, proportionality

02:18:30  23  is clearly a legal concept.

02:18:32  24      THE COURT:  It is.

02:18:32  25      MR. MOGIN:  Move to strike that testimony.

| | | |
|---|---|---|
| 02:18:34 | 1 | MR. MAROVITZ:  Well -- |
| 02:18:36 | 2 | THE COURT:  I mean, it is.  That is a -- in fact, |
| 02:18:40 | 3 | most of the lawyers in the room don't even know what it is, |
| 02:18:42 | 4 | but it is literally a legal concept.  I don't think most of |
| 02:18:50 | 5 | the vendors are thinking that. |
| 02:18:52 | 6 | MR. MAROVITZ:  They may be thinking about it a |
| 02:18:54 | 7 | different way. |
| 02:18:54 | 8 | THE COURT:  I have a question -- okay.  Your |
| 02:18:56 | 9 | objection is sustained. |
| 02:18:58 | 10 | Okay.  Are you saying that Temple -- Nan Nolan is one |
| 02:19:06 | 11 | of your custodians that you had picked out.  Am I |
| 02:19:10 | 12 | understanding you that when you began, you pulled every |
| 02:19:12 | 13 | document of Nan Nolan in Temple's entire system before you |
| 02:19:18 | 14 | applied the search terms? |
| 02:19:20 | 15 | THE WITNESS:  Just to be clear, your Honor, I didn't |
| 02:19:24 | 16 | pull the documents myself.  My understanding is that when |
| 02:19:26 | 17 | Temple identified Nan Nolan as a custodian -- |
| 02:19:30 | 18 | THE COURT:  One of the key custodians, okay. |
| 02:19:32 | 19 | THE WITNESS:  -- Temple pulled that person's |
| 02:19:34 | 20 | information off of their computer, off the email sever, and |
| 02:19:40 | 21 | off of any storage areas they had on their servers. |
| 02:19:42 | 22 | THE COURT:  Okay.  And then -- so when they had all |
| 02:19:48 | 23 | Nan Nolan's documents, they then applied the key words to |
| 02:19:52 | 24 | those documents? |
| 02:19:54 | 25 | THE WITNESS:  Correct.  They collected the documents, |

02:19:56  1  they moved them out of the normal company systems so they had

02:20:00  2  a preserved set, and it was on that set that they applied the

02:20:04  3  search terms.

02:20:04  4          THE COURT:  Thank you.

02:20:06  5  BY MR. MAROVITZ:

02:20:08  6  Q.  And to give this some more detail on the great question

02:20:12  7  that the judge asked you about, you know -- you may not know

02:20:16  8  each of the five custodians, but you know the basic

02:20:20  9  responsibilities that those custodians had with respect to

02:20:22  10  Temple-Inland that made part of that step?

02:20:26  11  A.  For Temple?

02:20:28  12  Q.  Yes.

02:20:28  13  A.  Key players?

02:20:28  14  Q.  Yes.

02:20:28  15  A.  Yes.  Those individuals had been described in the

02:20:32  16  correspondence.  I have seen that.  It wasn't just five

02:20:34  17  employees.  These were five senior-ranking executives within

02:20:40  18  the organization that had roles of president, vice president,

02:20:46  19  head of various divisions related to the corrugated box.

02:20:50  20  Q.  In fact, was Doyle Simons, the CEO, Pat Mayley, president,

02:20:54  21  Ron Zimbelman, Roy Lind, and Larry Norton all high-ranking

02:20:58  22  officials at Inland?

02:21:00  23  A.  I believe those are the correct names, yes.

02:21:02  24  Q.  Let me move on from proportionality.

02:21:06  25          How did each of the defendants here, each of the

| | | |
|---|---|---|
| 02:21:10 | 1 | defendants here, make a test for false negatives? |
| 02:21:12 | 2 | A.  Well, again, each of the -- well, first of all, |
| 02:21:20 | 3 | Georgia-Pacific, working with the other defendants, had |
| 02:21:22 | 4 | already tested the search terms not once but twice on their |
| 02:21:26 | 5 | own documents.  It was subsequent to that second testing that |
| 02:21:30 | 6 | the other six defendants took and customized the key words -- |
| 02:21:34 | 7 | the search strings, not key words, search strings for their |
| 02:21:38 | 8 | own organizations.  Each of those subsequent six again tested |
| 02:21:44 | 9 | -- executed the search terms against their documents, which |
| 02:21:46 | 10 | created a hit set and left behind a null set, and within the |
| 02:21:52 | 11 | null set drew forth a random sample of documents, between 500 |
| 02:21:58 | 12 | and 600 documents, and looked at those for potential |
| 02:22:04 | 13 | responsiveness. |
| 02:22:06 | 14 | MR. MOGIN:  Your Honor -- |
| 02:22:06 | 15 | THE WITNESS:  So they tested the null set. |
| 02:22:10 | 16 | MR. MOGIN:  -- it gets back to my preliminary |
| 02:22:12 | 17 | objection.  Is Mr. Regard testifying to the truth of the |
| 02:22:14 | 18 | matter of these facts about which he has no percipient |
| 02:22:18 | 19 | knowledge and the sole source of which is hearsay evidence or |
| 02:22:22 | 20 | evidence to way the plaintiffs have not been given access, |
| 02:22:26 | 21 | such as -- |
| 02:22:26 | 22 | THE COURT:  Such as what? |
| 02:22:28 | 23 | MR. MOGIN:  Such as discussions with the vendors, his |
| 02:22:30 | 24 | discussions with defense counsel that he's testified that he |
| 02:22:36 | 25 | relied upon in forming his opinions.  We have nothing. |

02:22:38  1        MR. MAROVITZ:  Judge, it simply isn't true they have

02:22:44  2   nothing.  We provided all of the documents.  We provided the

02:22:46  3   information that's contained on the chart, in the chart.  We

02:22:50  4   provided all the letters that we have now put together for

02:22:54  5   your Honor.

02:22:56  6        MR. MOGIN:  Let me be clear, your Honor, because this

02:22:58  7   is just --

02:22:58  8        MR. MAROVITZ:  What this is, Judge --

02:22:58  9        MR. MOGIN:  This is just not so, what Mr. Marovitz is

02:23:02  10  saying.  It is just not so.

02:23:02  11       THE COURT:  Well, this is pretty easy to figure out.

02:23:06  12       MR. MOGIN:  It certainly is.

02:23:06  13       THE COURT:  I mean, it's pretty easy to figure out.

02:23:08  14       MR. MOGIN:  When he says he provided us with the

02:23:10  15  information, he is saying that so he doesn't have to tell you

02:23:14  16  that he didn't provide us with any documents.  The information

02:23:20  17  is solely the information that's contained in their March 21st

02:23:24  18  letters.  That's it.  And as we have demonstrated, those

02:23:28  19  letters are pure hearsay.  We do not have, for example,

02:23:32  20  Mr. Regard's hit count.  We don't have that.

02:23:36  21       All the stuff with respect to the bottom portion of

02:23:40  22  the chart, which really is the guts of the chart and it's the

02:23:44  23  guts of Mr. Regard's testimony, we do not have.  And you

02:23:48  24  have -- in the letter from Temple-Inland that was submitted on

02:23:52  25  March 21st, you have their explanation.  They are not going to

02:23:58   1    give it to us supposedly because it either doesn't exist or

02:24:00   2    because it's privileged. If it's privileged and Mr. Regard

02:24:04   3    testifies to it, that can't possibly be admissible.

02:24:06   4            MR. MAROVITZ: Mr. Regard -- if I can finish,

02:24:08   5    Mr. Regard is not testifying as to any privileged information.

02:24:12   6    The information that's contained in Defense Exhibit 4 is

02:24:16   7    information that was based upon information that was provided

02:24:22   8    by the individual companies. Mr. Regard described the

02:24:24   9    conversations that he's had with those companies. We then

02:24:30  10    went ahead and documented all of the information that

02:24:34  11    Mr. Mogin and the plaintiffs asked for in the letters. That's

02:24:36  12    why we were here last time and why I started this examination.

02:24:40  13    And we went ahead and provided all this information.

02:24:46  14            There is no cache of documents that's being withheld

02:24:48  15    here. Mr. Regard was not funneled a lot of documents as part

02:24:52  16    of his interview process. He asked questions, he was given

02:24:56  17    answers. The documents that were created as part of the

02:25:00  18    letters to provide this information have been provided.

02:25:04  19            And one other thing, Mr. Regard is an expert witness.

02:25:08  20    I don't understand the whole point about this being hearsay.

02:25:10  21    He is providing the information upon which he is relying.

02:25:16  22            THE COURT: Well, one specific thing I do want to

02:25:18  23    talk about, and I think it came in a letter from you,

02:25:22  24    Mr. Marovitz, is I thought Mr. Mogin had said, We want the

02:25:28  25    null set materials. I don't know whether I am calling them

02:25:32  1  the right thing because I have never seen that null set.  I

02:25:36  2  don't know if you actually physically had a copy of the

02:25:42  3  documents that weren't responsive.  I mean, you must.

02:25:46  4  Somebody must have what was not responsive.  And that's the

02:25:50  5  only document, Mr. Mogin, that I understand they did not give

02:25:58  6  you.

02:26:00  7        MR. MOGIN:  What we don't have, your Honor, is, if

02:26:04  8  you recall, Mr. Regard testified that he pulled together

02:26:08  9  documents and pulled together information from various

02:26:10  10  sources.  Included in the sources were vendors who actually

02:26:14  11  performed the testing, the validation testing.

02:26:20  12        We have nothing as far as that goes.  We don't have a

02:26:24  13  single document, for example, that shows the hit count, just

02:26:30  14  the hit count.  We don't have a single document.  What we have

02:26:34  15  are letters from the defendants --

02:26:36  16        THE COURT:  Explaining what they did.

02:26:40  17        MR. MOGIN:  Right.  And if there is a better

02:26:44  18  illustration as to why those letters are not reliable in this

02:26:46  19  proceeding, it's the very fact that Mr. Neuwirth conceded this

02:26:52  20  morning that there was a mistake that was made, or at least a

02:26:56  21  distinction, between what he put on behalf of Georgia-Pacific

02:27:00  22  in the November 22nd letter and what his witness, Mr. Brown,

02:27:06  23  testified to, at least with respect to domain names.

02:27:10  24        Now, you can say, Oh, well, the domain names are a

02:27:12  25  small point.  But the fact of the matter is they're asking us

02:27:16    1    to rely upon lawyer letters, unsworn lawyer letters, which

02:27:22    2    they then turn over to the expert, which the expert then is

02:27:26    3    trying to bring in through the back door as facts.  It's

02:27:28    4    hearsay inside of hearsay inside of hearsay, none of which is

02:27:32    5    admissible.

02:27:34    6            Now -- and this gets back to some of the other

02:27:36    7    problems with Mr. Regard's testimony.  Let's assume for the

02:27:40    8    sake of argument, I am not conceding this point, assume for

02:27:44    9    the sake of argument that best practices has some relevance in

02:27:48    10   this proceeding.  I don't believe that it does.  Does it mean

02:27:52    11   that if they complied with best practices, that they don't

02:27:54    12   have to do anything more with respect to search?  What does

02:27:58    13   best practices have to do with anything?

02:28:00    14           The real issue here is did the defendants comply or

02:28:06    15   are they likely to comply with our request for production of

02:28:10    16   documents and is their method sufficient?  Whether it meets

02:28:14    17   best practices or not, as Mr. Regard has defined it, is

02:28:18    18   irrelevant.

02:28:20    19           The issue is, what's going to be the end result?  Are

02:28:24    20   we going to spend the next three years coming before your

02:28:26    21   Honor every 60 days to argue motions to compel because we have

02:28:32    22   discovered something more in the case?  Is that going to

02:28:34    23   happen?  Are the motions to compel going to be this high,

02:28:38    24   higher than even myself, your Honor?  Not that that's a

02:28:44    25   particular high bar, but you understand the point.

| | | |
|---|---|---|
| 02:28:48 | 1 | MR. MAROVITZ:  Your Honor, a few things. |
| 02:28:48 | 2 | First, Mr. Regard has already testified today there |
| 02:28:52 | 3 | is no cache of documents to which he has been provided that |
| 02:28:56 | 4 | Mr. Mogin doesn't have.  So let's put that behind us. |
| 02:28:58 | 5 | MR. MOGIN:  If I -- |
| 02:29:00 | 6 | MR. MAROVITZ:  If I could finish.  Second, if |
| 02:29:02 | 7 | Mr. Mogin wants to argue the case, there is an appropriate |
| 02:29:04 | 8 | time and place for that, and it's not during the middle of |
| 02:29:06 | 9 | witness testimony. |
| 02:29:08 | 10 | Third, the fact is that each of these defendants, if |
| 02:29:12 | 11 | your Honor believes it's important, could present declarations |
| 02:29:14 | 12 | or witnesses to walk through each of these.  That entire |
| 02:29:18 | 13 | process would take days, if not weeks.  I have never been |
| 02:29:22 | 14 | involved in a case like this before where that has been |
| 02:29:26 | 15 | required. |
| 02:29:26 | 16 | We have tried to streamline this in a way to provide |
| 02:29:30 | 17 | the court with a sufficient degree of education so that it |
| 02:29:32 | 18 | could make a reasoned, informed judgment, assuming, as |
| 02:29:38 | 19 | Dr. Lewis said before, that attorneys were operating in good |
| 02:29:42 | 20 | faith.  But we're happy, if the court believes that we need to |
| 02:29:44 | 21 | put in declarations or something like that, we will do that. |
| 02:29:46 | 22 | THE COURT:  I understand -- |
| 02:29:48 | 23 | MR. MAROVITZ:  I would just like to get -- if I may, |
| 02:29:50 | 24 | your Honor, I appreciate Mr. Mogin is protecting his record. |
| 02:29:52 | 25 | I understand that.  But I am very concerned about getting |

02:29:56   1   through Mr. Regard's testimony today.

02:29:56   2           THE COURT:  So am I.

02:29:58   3           MR. MAROVITZ:  And I'd like to be able to do that.

02:30:00   4           THE COURT:  Mr. Mogin, I am going to -- because I

02:30:02   5   have some idea on this case too.  You don't know the brilliant

02:30:10   6   ideas I have on the case.  I would like to hear what

02:30:12   7   Mr. Regard says.  I can do a number of things with his

02:30:16   8   testimony, I am fully aware of what I can do with it, but I am

02:30:22   9   going to -- you have a continuing objection.  I am going --

02:30:26  10   the record is clear.  You have a continuing objection on a

02:30:32  11   number of levels.  I think the record is clear.  Your motion

02:30:34  12   yesterday, you preserved your motion, and I would like to get

02:30:44  13   more information about these other six people.

02:30:48  14           I mean, this is the closest I've got to at least even

02:30:52  15   know what to ask as the next question.  Whether I need the six

02:30:56  16   people to come in here, whether I need an affidavit, I mean, I

02:31:00  17   do.  So I at least would like to hear what the fellow has got

02:31:04  18   to say here.

02:31:06  19           MR. MOGIN:  Very well, your Honor.  You had told me

02:31:08  20   to jump up, so I was jumping up.

02:31:10  21           THE COURT:  And I am the boss.  Okay.  Bye.

02:31:12  22           MR. MAROVITZ:  Thank you, your Honor.

02:31:12  23   BY MR. MAROVITZ:

02:31:14  24   Q.  Mr. Regard, I see that Judge Nolan has before her

02:31:16  25   Exhibit 4.  That's tab D, I think, to your report.  If you

02:31:20   1   could open that up, maybe that's a good place to start.

02:31:24   2   A.  Yes, sir.

02:31:26   3   Q.  Tell us what Exhibit 4 is.

02:31:26   4   A.  Well, as I described earlier, in the course of my

02:31:34   5   engagement, I reviewed the historical correspondence and

02:31:38   6   pleadings in this case specifically around the selection of

02:31:42   7   key players and the development of search terms, but I was

02:31:46   8   also given the opportunity to have telephonic conversations

02:31:50   9   with the various defendants.  In preparation and in the course

02:31:54   10  of that discussion, your Honor, I realized I was going to have

02:32:00   11  a lot of information I wanted to track, and I wanted to make

02:32:02   12  sure I was focused on information I was asking, so I created

02:32:06   13  this chart to facilitate those discussions and to help me

02:32:10   14  organize the information that I was learning.  This was

02:32:12   15  originally created as a way to help me recall these

02:32:14   16  conversations.

02:32:18   17        And so I created this chart, which has one column for

02:32:22   18  each of the six defendants I spoke with, and the first column

02:32:26   19  is a list of textual descriptions of the input for each of the

02:32:30   20  other six columns.  So that's the high-level review of this

02:32:34   21  chart.

02:32:34   22  Q.  So walk us through the chart so we understand what it

02:32:38   23  means.

02:32:38   24  A.  Well, I will go line by line.  The first line is the

02:32:44   25  defendant.  For each of the columns, I listed the name of the

| | | |
|---|---|---|
| 02:32:46 | 1 | defendant for whom this information pertained. |
| 02:32:50 | 2 | The next line is counsel. For each of the |
| 02:32:52 | 3 | defendants, I listed the name of the counsel that I gathered |
| 02:32:54 | 4 | or that I learned through my conversations. |
| 02:32:58 | 5 | The next line is the technology consultants. I had a |
| 02:33:02 | 6 | chance to speak with the consultants for each one of the |
| 02:33:06 | 7 | defendants, and I recorded their organization here. The |
| 02:33:08 | 8 | documents were reviewed by. One of my questions was, who did |
| 02:33:12 | 9 | the document review for the testing purposes. And I recorded |
| 02:33:16 | 10 | here that it was in the -- for example, in the case of |
| 02:33:20 | 11 | Temple-Inland, it was Mayer Brown that did the document |
| 02:33:22 | 12 | review. |
| 02:33:24 | 13 | The next line, the processing and search term |
| 02:33:28 | 14 | technology. The defendants did not use the same software. |
| 02:33:32 | 15 | They didn't use the same consultant or vendor. Here I |
| 02:33:36 | 16 | recorded the various software tools that were used by the |
| 02:33:38 | 17 | defendants for this process of gathering -- not gathering, but |
| 02:33:42 | 18 | indexing documents, managing documents and providing search |
| 02:33:50 | 19 | technology and the other technologies that were used. |
| 02:33:52 | 20 | The number of custodians. Here, I marked down -- |
| 02:33:56 | 21 | again, originally, your Honor, this originally was to help me |
| 02:34:00 | 22 | remember all these conversations. |
| 02:34:00 | 23 | THE COURT: Okay. |
| 02:34:02 | 24 | THE WITNESS: It only became later the thought that |
| 02:34:04 | 25 | this would be helpful in this court as well. |

02:34:08  1    I marked down the number of custodians that had been

02:34:10  2  collected.  Each of these marks it down.  The only difference

02:34:14  3  really is Temple-Inland, which has a stroke, 28-stroke-five,

02:34:20  4  because I learned that they used five custodians for their

02:34:22  5  initial key word search testing but ultimately gathered 28

02:34:28  6  custodians and used that for their document collection

02:34:32  7  processing and review.

02:34:34  8    Search terms used to collect documents.  I wanted to

02:34:38  9  know if search terms were used to collect documents from the

02:34:40  10  organization or if they were collected, as we discussed,

02:34:44  11  wholesale.  I learned that none of the defendants used search

02:34:46  12  terms to collect documents.  They collected them, which really

02:34:50  13  facilitates the subsequent review of the null set, because you

02:34:54  14  now have the documents in one sandbox, as it were, so you can

02:34:58  15  find out which ones are hits and which ones are null rather

02:35:02  16  than going back into the native systems, which is very

02:35:04  17  difficult.

02:35:04  18    The next item is modifications disclosed to

02:35:12  19  plaintiffs.  Part of best practices is to discuss with a

02:35:16  20  counter-party what the process is that you're going through.

02:35:18  21  Not everybody does that, but it certainly is something that's

02:35:22  22  highly recommended.  Here, each one of the defendants

02:35:24  23  disclosed the search terms that they were applying themselves,

02:35:28  24  and I marked down the dates that they were disclosed.

02:35:32  25    I took some notes in the current course of my

| | | |
|---|---|---|
| 02:35:34 | 1 | conversations.  I wrote these down on the chart.  I can go |
| 02:35:38 | 2 | through those, your Honor, or I can just skip down to the |
| 02:35:40 | 3 | bottom. |
| 02:35:44 | 4 | BY MR. MAROVITZ: |
| 02:35:44 | 5 | Q.  Mr. Regard, unless -- |
| 02:35:44 | 6 | THE COURT:  No, I don't. |
| 02:35:44 | 7 | BY MR. MAROVITZ: |
| 02:35:46 | 8 | Q.  Please skip down to the bottom for us. |
| 02:35:48 | 9 | A.  Down at the bottom again, I marked the names of the |
| 02:35:50 | 10 | defendants, how many documents they had collected for testing |
| 02:35:56 | 11 | purposes, especially the case of Temple-Inland. |
| 02:35:58 | 12 | THE COURT:  All right.  So where did you get that |
| 02:36:02 | 13 | information? |
| 02:36:02 | 14 | THE WITNESS:  When I was on the phone, I asked for |
| 02:36:04 | 15 | it. |
| 02:36:04 | 16 | THE COURT:  Well, I know, but, I mean, did -- when |
| 02:36:08 | 17 | you're saying you talked to the defendant, I mean, you must |
| 02:36:12 | 18 | have talked to -- okay.  So for Temple-Inland -- |
| 02:36:16 | 19 | THE WITNESS:  I was on the phone. |
| 02:36:18 | 20 | THE COURT:   -- who or what type of person told you |
| 02:36:20 | 21 | about the collection?  That's what I want to know.  Or did you |
| 02:36:26 | 22 | have a printout on this?  How did you get 183,757, this |
| 02:36:34 | 23 | number? |
| 02:36:34 | 24 | THE WITNESS:  I was on the phone with counsel and the |
| 02:36:38 | 25 | consultants for Temple-Inland -- |

| | | |
|---|---|---|
| 02:36:38 | 1 | THE COURT: Okay. |
| 02:36:40 | 2 | THE WITNESS: -- and I asked them, do they know how |
| 02:36:42 | 3 | many documents they collected. And they said, Yes, we do. |
| 02:36:46 | 4 | THE COURT: All right. |
| 02:36:46 | 5 | THE WITNESS: The number is, and they gave me the |
| 02:36:48 | 6 | number 183,757, and I recorded it. |
| 02:36:54 | 7 | BY MR. MAROVITZ: |
| 02:36:56 | 8 | Q. Is the same thing true, Mr. Regard, for hits, null set, |
| 02:36:58 | 9 | test set, and test set responsive? Why don't you walk us |
| 02:37:04 | 10 | through where that came from and how that works. |
| 02:37:08 | 11 | A. I asked them how many documents they had collected, how |
| 02:37:10 | 12 | many hits they had when they ran the search terms. I don't |
| 02:37:12 | 13 | recall if I asked them how large the null set was or if I did |
| 02:37:16 | 14 | the math myself, but you will note that the null set plus the |
| 02:37:18 | 15 | hits is equal to the collection, which is what I would expect. |
| 02:37:22 | 16 | I asked them if they had tested documents and, if so, how many |
| 02:37:26 | 17 | they had tested randomly. |
| 02:37:30 | 18 | These are the answers I recorded under the test set. |
| 02:37:32 | 19 | I asked how many documents had been found to be responsive. I |
| 02:37:36 | 20 | recorded that down. And either they had reported to me or I |
| 02:37:42 | 21 | did the math again. In this case of Temple-Inland, the first |
| 02:37:46 | 22 | column, 1.4 percent is the result of dividing 7 by 500. |
| 02:37:54 | 23 | Q. Now, Mr. Regard, your chart at the bottom in the left, it |
| 02:38:00 | 24 | says, The defendants are 95 percent confident that the null |
| 02:38:04 | 25 | set contains no more than this percent of potentially relevant |

| | | |
|---|---|---|
| 02:38:08 | 1 | documents, plus or minus 5 percent or less. |
| 02:38:10 | 2 | What did you mean there by "potentially relevant |
| 02:38:14 | 3 | documents"? |
| 02:38:14 | 4 | A.  Well, that whole phrase is sort of the -- I want to say |
| 02:38:20 | 5 | the worst case scenario, it is the most conservative |
| 02:38:24 | 6 | interpretation of the various testings that all the defendants |
| 02:38:30 | 7 | did. |
| 02:38:30 | 8 | When I say potentially relevant documents, I'm |
| 02:38:32 | 9 | referring to documents relevant to this testing, documents |
| 02:38:34 | 10 | that would have been responsive to the search -- not to the |
| 02:38:38 | 11 | search terms, because they weren't responsive to the search |
| 02:38:42 | 12 | terms, but that the defendants found when they tested the null |
| 02:38:44 | 13 | set would have otherwise have been responsive to the document |
| 02:38:48 | 14 | requests from the plaintiffs. |
| 02:38:48 | 15 | Q.  Mr. Regard, does Defense Exhibit 4 fairly and accurately |
| 02:38:54 | 16 | summarize the information you obtained in your investigation |
| 02:38:56 | 17 | and study of document review procedures? |
| 02:39:00 | 18 | A.  It reflects it precisely. |
| 02:39:06 | 19 | MR. MAROVITZ:  I understand, your Honor, this is |
| 02:39:06 | 20 | subject to Mr. Mogin's objection and to your Honor's decision, |
| 02:39:10 | 21 | obviously, but we would move for admission of the report. |
| 02:39:14 | 22 | THE COURT:  I am certainly going to take that under |
| 02:39:16 | 23 | advisement. |
| 02:39:16 | 24 | BY MR. MAROVITZ: |
| 02:39:20 | 25 | Q.  By the way, Mr. Regard, do you understand the defendants |

02:39:22  1  generate random samples?

02:39:24  2  A.  Yes, sir.

02:39:26  3  Q.  And are those samples generated in different ways by

02:39:32  4  different processing and search term technology platforms?

02:39:38  5  A.  They are.  Every platform that I am familiar with has the

02:39:42  6  ability to generate random samples, but as Dr. Lewis

02:39:46  7  testified, that can also be done external to the technology

02:39:50  8  itself.

02:39:50  9  Q.  Have you seen any literature in which you have seen that

02:39:56  10  there is a suggestion, for instance, Clearwell's randomizer,

02:40:02  11  features bias in some way?

02:40:04  12  A.  I have not seen anything like that.

02:40:06  13  Q.  What did the testing tell you as an experienced

02:40:08  14  practitioner who is familiar with best practices about the

02:40:12  15  performance of the ESI protocol used here?

02:40:16  16         THE COURT:  Well, as to what -- as to which part?  I

02:40:22  17  mean, we now all know there's a number of steps.  And so as

02:40:28  18  to --

02:40:32  19         MR. MAROVITZ:  Fair point, your Honor.  What I was

02:40:34  20  driving at was all the steps taken together as part of the ESI

02:40:38  21  protocol, Mr. Regard testified earlier that he thought that

02:40:40  22  defendants' process met or exceeded best practices, and I was

02:40:44  23  trying to bring to circle all the way from the end.

02:40:48  24         So I was curious after he has been through this

02:40:50  25  process, asked these questions to the defendants and to their

02:40:56   1   vendors, their consultants, what conclusions he reached based

02:41:00   2   upon all the data that he was given as an expert witness in

02:41:04   3   this case.

02:41:18   4           THE COURT:  Well, this is really hard to do because

02:41:30   5   we know, or at least I feel like I know about Georgia-Pacific,

02:41:38   6   how many random samples they took, what they did after they

02:41:42   7   got the random samples, and you're asking him a very general

02:41:44   8   question here where we have none of that subpart here at all.

02:41:50   9   He didn't perform any of the subpart.

02:41:56   10          So I don't know what he would know.  I don't know

02:42:00   11  what he would know -- I mean, I think he can tell us other

02:42:06   12  things.  I don't know what that question is -- how that

02:42:10   13  question is going to move the ball here.

02:42:12   14          MR. MAROVITZ:  Let me ask it a different way, if I

02:42:14   15  may, Judge?

02:42:14   16          THE COURT:  All right.

02:42:16   17  BY MR. MAROVITZ:

02:42:16   18  Q.  Just looking at the results that were obtained, you can

02:42:18   19  see, Mr. Regard, that the results that were obtained were

02:42:22   20  different for each of the six custodians, correct?

02:42:26   21  A.  Six defendants?

02:42:28   22  Q.  Fair enough, six defendants.

02:42:30   23  A.  Yes, sir.

02:42:30   24  Q.  And, obviously, this chart -- as Judge Nolan has pointed

02:42:34   25  out, this chart does not include Georgia-Pacific, so it's the

02:42:38  1  other six defendants.

02:42:40  2          What does the fact that these results are different

02:42:44  3  -- so you didn't perform these, but they were given to you.

02:42:46  4  What does the fact that these results are different tell you,

02:42:50  5  in your expert opinion?

02:42:52  6  A.  I think it would be natural to question whether or not

02:42:54  7  they would have been identical because each of the defendants

02:42:56  8  had different document sets, each of them used different

02:43:00  9  technologies, each one of them modified the search terms to

02:43:04  10  fit their personal technology and their personal organization.

02:43:10  11  And as I testified, each one of them relied on the initial

02:43:14  12  search terms that were developed through the iterative process

02:43:18  13  that Georgia-Pacific went through.  Not once, but

02:43:20  14  Georgia-Pacific went through it twice.  Again, they had input

02:43:24  15  to it, but it was only tried against Georgia-Pacific's

02:43:26  16  documents.

02:43:28  17          So the third round of application of search terms by

02:43:32  18  each of the defendants individually is where I would have an

02:43:36  19  interest in seeing how those search terms performed.

02:43:40  20          And these different numbers tell me two things.  One

02:43:48  21  is that, at least in this null set measuring, if you accept

02:43:50  22  these numbers, that the residual percentage of responsive

02:43:56  23  documents was very low; but more importantly, from a

02:44:02  24  confirmation of the process, that the variation between the

02:44:06  25  different defendants is very narrow.  That within the narrow

| | | |
|---|---|---|
| 02:44:10 | 1 | band of variance, they all had less than a 5 percent residual |
| 02:44:14 | 2 | rate of potentially responsive documents. |
| 02:44:18 | 3 | So it doesn't make the process a reasonable |
| 02:44:22 | 4 | thought-out process, but it does confirm that the reasonable |
| 02:44:26 | 5 | thought-out process that was ascribed to me did, in fact, |
| 02:44:30 | 6 | produce good results. |
| 02:44:32 | 7 | Q. I'd like to focus on a different topic, if we can, and |
| 02:44:36 | 8 | that's particularly -- unless, Judge, you have any other |
| 02:44:38 | 9 | questions? |
| 02:44:38 | 10 | THE COURT: No. |
| 02:44:38 | 11 | BY MR. MAROVITZ: |
| 02:44:40 | 12 | Q. -- (continuing) particularly on advantages to properly use |
| 02:44:44 | 13 | key words. Okay? |
| 02:44:44 | 14 | A. Okay. |
| 02:44:46 | 15 | Q. While you certainly have to look at the specifics of any |
| 02:44:48 | 16 | ESI collection search procedure, whether it's Boolean |
| 02:44:54 | 17 | searching, some form of predictive coding, some combination of |
| 02:44:58 | 18 | the two, or something entirely different, do you believe, in |
| 02:45:02 | 19 | your expert opinion, that Boolean key words present some |
| 02:45:06 | 20 | inherent advantages? |
| 02:45:06 | 21 | A. Like any system, there are advantages -- every system has |
| 02:45:08 | 22 | its own advantages. Boolean searching certainly has its |
| 02:45:12 | 23 | advantages. |
| 02:45:12 | 24 | Q. And let's walk through some of those. |
| 02:45:18 | 25 | First, do you believe that there are transparency |

02:45:20   1   advantages to Boolean searching?

02:45:22   2   A.  I do.

02:45:24   3   Q.  How so?

02:45:24   4   A.  Well, when we work with Boolean searching, and as Boolean

02:45:30   5   terms were used in the documentation in this case, they can

02:45:34   6   actually be printed out, they can be disclosed, they can be

02:45:36   7   discussed.  The idea that you can actually write them out

02:45:42   8   gives them that degree of transparency; you can see them, you

02:45:46   9   can work with them, you can see what the computer is using to

02:45:48  10   apply those search terms.

02:45:50  11   Q.  And how about familiarity?  Does Boolean searching provide

02:45:56  12   some familiarity to the process?

02:45:58  13   A.  I think it's one of the things we are talking about today

02:46:00  14   is, are we talking about existing technology versus emerging

02:46:04  15   technologies.

02:46:04  16         Boolean searching has been around for a very long

02:46:08  17   time.  We had searches going back -- the testimony this

02:46:12  18   morning, Blair and Maron study goes back to 1985, they were

02:46:16  19   already looking at Boolean search technology.  Thanks to the

02:46:22  20   Internet, it's something people are very familiar with.

02:46:24  21   Q.  How about proximity; does Boolean searching provide any

02:46:28  22   kind of advantage when it comes to proximity?

02:46:30  23   A.  Yes, it does.  The Boolean search tools that we use today

02:46:34  24   allow us to specify word proximity.  The ability to require

02:46:42  25   that two words be within a phrase, used in quotations, the

| | | |
|---|---|---|
| 02:46:48 | 1 | ability to require that two words be within a certain number |
| 02:46:50 | 2 | of words of each other is helpful when trying to limit the |
| 02:46:56 | 3 | number of documents and focus documents on key phrases rather |
| 02:47:00 | 4 | than that those words would appear anywhere in the document. |
| 02:47:04 | 5 | Q.  Does the Boolean approach present advantages for |
| 02:47:08 | 6 | commutative properties? |
| 02:47:10 | 7 | A.  Again, I feel that it does in that Boolean search terms, I |
| 02:47:14 | 8 | find, are not commutative sensitive.  Obviously, a Boolean |
| 02:47:22 | 9 | mathematical string has commutative issues with ands and ors, |
| 02:47:26 | 10 | but once you have the string and you apply it to a set of |
| 02:47:30 | 11 | documents, you will get the same results whether you apply |
| 02:47:32 | 12 | that to documents 1 through 100 or documents 100 through 1. |
| 02:47:38 | 13 | It doesn't matter which order you apply the documents.  I |
| 02:47:40 | 14 | don't see that same commutative response with every type of |
| 02:47:44 | 15 | information retrieval system. |
| 02:47:46 | 16 | Q.  And you heard Dr. Lewis testify this morning about how |
| 02:47:50 | 17 | machine learning might produce different results based upon |
| 02:47:56 | 18 | different software that was used for different viewers. |
| 02:47:58 | 19 | Does a Boolean approach have advantages with respect |
| 02:48:02 | 20 | to repeatability? |
| 02:48:04 | 21 | A.  We tend to find that Boolean search strings execute |
| 02:48:06 | 22 | similarly, if not identically, across multiple platforms, |
| 02:48:10 | 23 | competing software tools.  So it's been my experience that |
| 02:48:16 | 24 | when I describe a search that I ran on a set of documents in |
| 02:48:20 | 25 | one software tool to someone and they execute it on the same |

02:48:24  1  documents on a different software tool, they get the same

02:48:26  2  results.

02:48:26  3  Q.  You listed a number of advantages to Boolean searches.

02:48:30  4  Are there any advantages to machine learning that you are

02:48:32  5  aware of?

02:48:34  6  A.  Absolutely.

02:48:36  7  Q.  What's one of those, for example?

02:48:36  8  A.  One of the advantages of machine learning is that if you

02:48:44  9  are willing to rely on the training set and rely on the system

02:48:48  10  that you're using that you're training, then you can curtail

02:48:52  11  the number of documents you need to look at.  There are

02:48:56  12  tradeoffs for that, but you can -- in the interest of time, if

02:48:58  13  you're interested in the amount of time that it takes to

02:49:00  14  review documents, studies suggested that you can look at fewer

02:49:06  15  documents in a shorter amount of time using properly trained

02:49:12  16  machine learning systems.

02:49:12  17  Q.  You said, "There are tradeoffs for that."  What are some

02:49:14  18  of those?

02:49:16  19  A.  Well, machine learning is still expensive.  It's an

02:49:22  20  expensive process.  You need to still pay to have your

02:49:24  21  documents indexed, and the systems that I am familiar with

02:49:28  22  have a surcharge for machine learning indexing and technology

02:49:32  23  versus indexing for search terms.

02:49:34  24  Q.  Are there any others that you can think of in this

02:49:38  25  context?

02:49:38   1   A.  Well, I think that Dr. Lewis has described at great length

02:49:46   2   the amount of care necessary for the proper training of

02:49:48   3   machine learning systems.

02:49:50   4         You also -- when you use machine learning, you don't

02:49:52   5   need to go through the process of developing the key words,

02:49:56   6   though, that were done in this process.  So there is a

02:49:58   7   tradeoff there.  You spend more time perhaps on the training

02:50:02   8   but less time developing the key words.

02:50:04   9         In this case, the defendants in Georgia-Pacific spent

02:50:08   10  a lot of time developing key words.

02:50:10   11  Q.  Based upon all of this information that you received

02:50:14   12  during your conversations with the defendants, the

02:50:18   13  correspondence that you were given, based upon your expertise

02:50:20   14  and your knowledge as an expert in this particular industry,

02:50:24   15  do you have an opinion on whether the defendants' search

02:50:26   16  protocol, which included but was not limited to key words, met

02:50:30   17  or exceeded best practices?

02:50:32   18  A.  Well, you said included but not limited to key words.  We

02:50:36   19  didn't really discuss that.  I believe it's in my report.

02:50:40   20         But overall, in general, I find that the process

02:50:44   21  followed by the defendants is the same or exceeds the process

02:50:50   22  that I recommend to my own clients in cases that I have worked

02:50:52   23  on, and it's consistent with the work that I have done on best

02:50:56   24  practices at the Sedona Conference and elsewhere.

02:51:00   25  Q.  I don't want to cut you short on any part of something

| | | |
|---|---|---|
| 02:51:04 | 1 | that's important for the court to understand.  You said |
| 02:51:06 | 2 | there's something in your report about the fact that this |
| 02:51:08 | 3 | particular protocol relied upon key words but was not limited |
| 02:51:10 | 4 | to them, and we haven't discussed it.  What did you mean by |
| 02:51:16 | 5 | that? |
| 02:51:16 | 6 | A.  In the course of developing the key words, the defendants |
| 02:51:20 | 7 | also, and they disclosed this, have relied upon other -- not |
| 02:51:28 | 8 | machine learning, but the linguistic functionality of |
| 02:51:32 | 9 | Clearwell and the tools that they're using to test the null |
| 02:51:34 | 10 | set, to look at the residual documents, to use context |
| 02:51:38 | 11 | searching or topic searching, to use other methodologies to |
| 02:51:42 | 12 | see are there indices, that there are sets of identifiable |
| 02:51:48 | 13 | patterns and documents that are being left behind. |
| 02:51:52 | 14 | THE COURT:  All right.  Did Georgia-Pacific, because |
| 02:51:54 | 15 | I don't have that testimony memorized, did they go into -- on |
| 02:52:02 | 16 | their direct or on their cross, did they go into that they |
| 02:52:10 | 17 | used computer-assisted materials? |
| 02:52:12 | 18 | MR. NEUWIRTH:  Yes, I believe that that was discussed |
| 02:52:14 | 19 | in some detail by Mr. Koch and by Mr. Brown, and if it would |
| 02:52:20 | 20 | be helpful to the court, after the hearing, we can direct you |
| 02:52:22 | 21 | to where that is in the transcript. |
| 02:52:26 | 22 | THE COURT:  All right. |
| 02:52:26 | 23 | Now, do you know, do you actually know, or did you |
| 02:52:32 | 24 | discuss with any of the other six here what kind of |
| 02:52:40 | 25 | computer-assisted machine, whatever you want to call it, each |

02:52:48  1  of the defendants used and at what stage they used it or how

02:52:54  2  they used it?  Do you know?

02:52:56  3          THE WITNESS:  I don't recall --

02:52:56  4          THE COURT:  With Temple.  Do you know, because you

02:52:58  5  were working for Temple?  Do you know what Temple did?

02:53:02  6          THE WITNESS:  Yes, ma'am.

02:53:02  7          THE COURT:  Okay.

02:53:04  8          THE WITNESS:  My recollection is after Temple ran the

02:53:08  9  search terms that they had modified for their organization,

02:53:14  10  they looked at the null set documents, they chose a random set

02:53:18  11  of the null set documents.

02:53:20  12          And then after they looked at the null set, they

02:53:22  13  also, using the Clearwell functionality of topics, looked at

02:53:30  14  the topics that Clearwell automatically generates associated

02:53:36  15  with the null set documents to see if there were phrases or

02:53:42  16  topics that seemed indicative of likely responsive documents.

02:53:48  17          THE COURT:  All right.  So then what did they do with

02:53:54  18  that information when they got it?

02:53:56  19          THE WITNESS:  Then they identified a few topics that

02:54:00  20  looked as if --

02:54:02  21          THE COURT:  Do you then make -- if you are using

02:54:04  22  Boolean, then do you turn what you learn from Clearwell into a

02:54:08  23  key word search?  Do you add more key words?  Is that what you

02:54:12  24  do?

02:54:12  25          THE WITNESS:  You can, your Honor.  Typically, what

02:54:14  1  would happen would be -- remember, we are looking at the end

02:54:18  2  of a testing process as opposed to the middle of the revisions

02:54:22  3  and modifications.  What would normally happen during

02:54:24  4  revisions and modifications is that you would run search terms

02:54:28  5  against your documents --

02:54:30  6       THE COURT:  Right.

02:54:30  7       THE WITNESS:  -- you would get some hits, you would

02:54:32  8  look at them and say, Some of these are actually responsive

02:54:36  9  and some of these are not.

02:54:38  10       THE COURT:  Okay.

02:54:38  11       THE WITNESS:  So if we are getting documents that are

02:54:40  12  not responsive, we probably want to change our search terms

02:54:44  13  and see if we can change them in a way that we continue to get

02:54:46  14  the ones we want and we don't get the ones we don't want.

02:54:50  15       And then you look at the documents you didn't get,

02:54:52  16  the null set, you look at some of those, and you say, We took

02:54:56  17  a sample of the null set.  Some of these were responsive, some

02:55:00  18  of these are not.  Why were these responsive documents not

02:55:02  19  responding to our search terms?  And you go back and

02:55:06  20  re-examine your search terms and decide, do you need to make

02:55:08  21  modifications and can you?

02:55:10  22       Sometimes you might see a word in a document you

02:55:14  23  missed and say, We should add that, but then when you add it,

02:55:18  24  it brings in 500 new documents that are totally nonresponsive.

02:55:22  25  So it's constantly a balancing test between am I getting more

02:55:26   1   good than bad when I make this change.

02:55:30   2   BY MR. MAROVITZ:

02:55:30   3   Q. Mr. Regard, just to be clear, the non-GP defendants were

02:55:36   4   able as part of the entire process that you have described

02:55:38   5   before to benefit from the GP iterative process that we have

02:55:44   6   heard about in the last hearing, correct?

02:55:46   7   A. Not just benefit. They participated in it. They gave

02:55:50   8   feedback, they gave commentary, and they helped create those

02:55:52   9   search terms.

02:55:52  10   Q. I want to make sure the record is clear on one thing. You

02:55:56  11   mentioned Clearwell a minute ago. On your chart, Defense

02:56:00  12   Exhibit 4, if you look at the processing of search term

02:56:04  13   technology, some of the defendants used Clearwell and some of

02:56:06  14   the defendants used other processing search term technology.

02:56:10  15        Is it your testimony that all of the defendants used

02:56:12  16   Clearwell?

02:56:14  17   A. No. I thought I had made that clear.

02:56:16  18   Q. I may have misheard you.

02:56:18  19   A. No, they did not all use Clearwell. I was talking

02:56:22  20   specifically about Temple.

02:56:22  21   Q. And one other small point. I think you had mentioned

02:56:24  22   earlier that there had been 15 search strings, one prepared

02:56:30  23   for each request for production. In fact, some of the search

02:56:34  24   strings that the defendants prepared actually covered multiple

02:56:38  25   requests for production, did they not?

02:56:38   1   A.   I believe that's correct.   There is not a one-to-one

02:56:46   2   correlation between the search strings and individual requests

02:56:48   3   for production.

02:56:48   4             MR. MAROVITZ:   Your Honor, if I could turn to

02:56:50   5   custodians, I'd be pleased to do that unless the court has

02:56:54   6   additional questions on this.

02:56:56   7             THE COURT:   Is this -- this came to our office.   What

02:57:02   8   does this --

02:57:06   9             MR. MAROVITZ:   That's not mine.

02:57:10  10             MS. MILLER:   That's plaintiffs', your Honor.

02:57:12  11             THE COURT:   That's plaintiffs'.

02:57:12  12             MR. MAROVITZ:   You may see that later.

02:57:14  13             THE COURT:   Okay.

02:57:16  14             MR. MAROVITZ:   Mr. Regard may see that later.

02:57:20  15             THE WITNESS:   Would you like to see it in chambers?

02:57:22  16             THE COURT:   Okay.

02:57:22  17             Yes, you can go on to custodians.

02:57:24  18   BY MR. MAROVITZ:

02:57:26  19   Q.   Mr. Regard, are you familiar with the method -- let me

02:57:28  20   start again.

02:57:28  21             I think you have already testified today about the

02:57:32  22   method that the defendants used in identifying custodians and

02:57:36  23   document sources in this case.

02:57:42  24             Just describe for us, in general terms so we can

02:57:48  25   understand ourselves, how that's done.

02:57:50  1  A.  Well, in this particular case, the defendants chose

02:57:52  2  custodians based on the document requests they had received

02:57:56  3  from plaintiffs and the allegations in the case.  And it's

02:58:00  4  already been put forth in the record, or at least in the

02:58:04  5  correspondence, that they chose key players who had roles and

02:58:08  6  responsibilities consistent with the documents requested in

02:58:14  7  the allegations and they chose key players who were expected

02:58:18  8  to have the highest concentration of responsive documents.

02:58:22  9  Q.  You understand that Mr. Hanners, an expert who has

02:58:26  10  testified on behalf of plaintiffs, has taken the position that

02:58:30  11  defendants should be required to collect fully and completely

02:58:32  12  all information on a corporate function department or work

02:58:36  13  unit basis instead of by custodian only, correct?

02:58:40  14  A.  I did hear that testimony, yes.

02:58:42  15  Q.  And according to Mr. Hanners, the failure to do that in

02:58:46  16  general investigative terms is like searching a house for

02:58:50  17  computer information but not looking into every room of the

02:58:52  18  house, correct?

02:58:54  19  A.  I heard the testimony.  I don't necessarily agree with it.

02:58:58  20  Q.  Let me ask you about that.  Without regard to whether a

02:59:00  21  party uses a Boolean search approach, a machine learning

02:59:04  22  approach, or some other approach, do you have an opinion on

02:59:08  23  the current prevailing best practice for the source of

02:59:12  24  collection of ESI?

02:59:12  25  A.  I know from my experience what my clients have done in

02:59:20 1  multiple litigations over the last 15, 20 years, and that has

02:59:24 2  been to collect documents on a custodial basis.  The reason

02:59:28 3  behind that is because documents, by and large, are organized

02:59:32 4  on a custodial basis within organizations.  We all use the

02:59:36 5  information technology ourselves.  When I go to collect

02:59:42 6  documents within organizations, we see documents organized by

02:59:46 7  mailbox, by desktop, by individual storage area on the

02:59:50 8  servers, and so when we collect information, we collect it

02:59:54 9  also on that same custodial basis.  And my understanding is,

02:59:58 10 by and large, although not entirely, that's the way the

03:00:02 11 information is collected here.

03:00:02 12       To go instead to every single custodian or to every

03:00:06 13 single server or to every single system when there is no

03:00:10 14 expectation of finding responsive documents there as

03:00:14 15 Mr. Hanners would suggest is simply not done.  It's not

03:00:18 16 appropriate.

03:00:20 17 Q.  And you said "by and large" a minute ago, custodian-based

03:00:26 18 approach by and large.  What did you mean by that?

03:00:28 19 A.  Well, we find areas of document storage where it came that

03:00:36 20 people may share information.  And my understanding is that

03:00:38 21 those sources were also considered, especially applications

03:00:40 22 that have information.  Whether it be financial or inventory,

03:00:44 23 that type of information is not really a person-custodian

03:00:50 24 specific; it's a function specific.  It's also been identified

03:00:54 25 and collected.

03:00:54  1  Q.  What's the harm in searching every room in the house,

03:00:58  2  searching everywhere documents are located, and then

03:01:02  3  separating the wheat from the chaff later on?

03:01:06  4  A.  There's a lot of problems with that approach.  Number one,

03:01:08  5  it's obviously time-consuming to go everywhere instead of

03:01:12  6  where you naturally think the documents are.  To use the

03:01:16  7  analogy of a house, if it's your house and you needed to go

03:01:18  8  find a book, you wouldn't go in every room.  You'd go to where

03:01:22  9  you keep your books.

03:01:22  10  And so it's a lot easier when companies understand

03:01:26  11  their own systems and their own organization to identify those

03:01:30  12  areas most likely to have the information that's being

03:01:32  13  requested.  I don't know if that's a best practice.  It's a

03:01:36  14  prevailing practice.  That's what I see in every case I work

03:01:38  15  in.

03:01:40  16  The downside of collecting everything and, to use

03:01:42  17  your phrase, separating the wheat from the chaff, is that's an

03:01:46  18  expensive process.  It's expensive and it's time-consuming.

03:01:50  19  And these collections of documents, the processing, the

03:01:54  20  indexing, even the application of machine learning, if that

03:01:58  21  had been someone's choice, is not free.  There is an expense

03:02:02  22  associated with that.  And if you can just collect more and

03:02:04  23  more documents, you are just driving up the expense and the

03:02:06  24  delay.

03:02:06  25  Q.  Let me turn to the third thing that you offered of the

| | | |
|---|---|---|
| 03:02:12 | 1 | four in this proceeding.  Have you reached an opinion as to |
| 03:02:14 | 2 | the organization and accessibility of defendants' ESI |
| 03:02:20 | 3 | productions? |
| 03:02:20 | 4 | A.  I have. |
| 03:02:22 | 5 | Q.  And with respect to that opinion, what exactly is it? |
| 03:02:26 | 6 | A.  My opinion is that the format of production that the |
| 03:02:34 | 7 | defendants have done or were in the process of doing is both |
| 03:02:38 | 8 | consistent with the stipulated order and it's consistent with |
| 03:02:42 | 9 | best practices on the production of ESI. |
| 03:02:44 | 10 | Q.  I want to get to the stipulated order in just a minute, |
| 03:02:48 | 11 | but first, let's talk about best practices with respect to |
| 03:02:52 | 12 | ESI.  What do you mean by best practices with respect to ESI |
| 03:02:56 | 13 | as it relates to the organization and accessibility? |
| 03:02:58 | 14 | A.  Well, best practices today doesn't relate specifically to |
| 03:03:06 | 15 | organization and accessibility; it relates to the manner in |
| 03:03:10 | 16 | which the information is provided and the richness of the |
| 03:03:14 | 17 | information that's provided. |
| 03:03:16 | 18 | In this litigation, producing documents by TIFF image |
| 03:03:20 | 19 | but with accompanying full text that was in the contents of |
| 03:03:24 | 20 | those documents and the metadata provides not only a |
| 03:03:30 | 21 | conveyance of those documents, both in photographic form, the |
| 03:03:34 | 22 | TIFF images, and the textual form; it provides anyone who |
| 03:03:38 | 23 | takes that production into a litigation management tool the |
| 03:03:42 | 24 | ability to conduct word searches, the ability to organize |
| 03:03:46 | 25 | those documents by any of the metadata fields, the ability to |

03:03:50  1  organize by chronology, by recipient, by sender, or any subset

03:03:54  2  thereof, to look at the documents because of the path, by the

03:04:00  3  way that they were organized in their original physical

03:04:04  4  locations of the companies.

03:04:06  5  So you end up providing the recipient, whoever the

03:04:08  6  recipient may be, with a great deal of functionality and how

03:04:12  7  those documents were organized.

03:04:14  8  Q.  And in terms of the way electronic files are actually kept

03:04:18  9  within a company, for those of us who are not --

03:04:20  10  THE COURT:  Did the plaintiffs ask for -- let's go

03:04:24  11  back to number 3.  Did the plaintiffs ask that the documents

03:04:30  12  be produced in any particular format that you know,

03:04:36  13  Mr. Regard?

03:04:36  14  THE WITNESS:  My understanding, your Honor, is that

03:04:38  15  they asked that the defendants would reorganize the documents

03:04:42  16  according to a list of 29 topics.

03:04:48  17  THE COURT:  That's not really the form of production,

03:04:52  18  though.

03:04:52  19  MR. MAROVITZ:  I might be able to help, your Honor,

03:04:54  20  with this question.

03:04:54  21  THE COURT:  I haven't been able to literally find

03:04:58  22  that in the thousands of pieces of paper.  So this is a lawyer

03:05:02  23  question as much as it's a Mr. Regard question.

03:05:04  24  MR. MOGIN:  I'm sorry, your Honor.  Are you asking

03:05:06  25  the physical format?  Did we request native format, for

03:05:10    1   example?  Is that what you are asking?

03:05:12    2              THE COURT:  Yes.  Did you ask specifically that

03:05:16    3   they -- first of all, I don't even know, have you received one

03:05:18    4   document in this case?  I mean, last night I thought to

03:05:24    5   myself, I don't even know if you have received any documents.

03:05:26    6   And if you did, did you receive the documents in a particular

03:05:30    7   format that you have requested?

03:05:32    8              MR. MOGIN:  Your Honor, we specified the format in

03:05:38    9   the request for production.

03:05:40   10              THE COURT:  Which was what?

03:05:40   11              MR. MOGIN:  It was -- just bear with me just a

03:05:44   12   second.

03:05:46   13              And then that was essentially superseded by the

03:05:50   14   production format order; that is, with respect to whether we

03:05:56   15   get TIFF images for certain things, native images for other

03:06:00   16   things, the metadata that was required to be produced, et

03:06:02   17   cetera.  Nothing, nothing, it's our position, in the ESI

03:06:10   18   format order signed by Judge Shadur --

03:06:12   19              THE COURT:  That's called agreed.

03:06:14   20              MR. MOGIN:  -- has anything whatsoever to do with the

03:06:20   21   organization of the documents.

03:06:26   22              THE COURT:  Is that the 29 categories?

03:06:28   23              MR. MOGIN:  It is.

03:06:28   24              THE COURT:  Okay.

03:06:30   25              MR. MOGIN:  And the 29 categories, and the need to

03:06:32 1    organize -- actually, not to organize, but to label, if you

03:06:38 2    will, or to somehow give us an indication of which documents

03:06:46 3    respond to which request, which was specifically laid out in

03:06:48 4    the production of documents.

03:06:50 5            MR. MAROVITZ:  Judge, if I can focus us on this

03:06:52 6    particular issue.

03:06:52 7            If I could hand up to Mr. Regard the stipulation that

03:06:56 8    was just described, along with the order of Judge Shadur.

03:07:00 9            THE COURT:  Sure.

03:07:00 10           MR. NEUWIRTH:  Your Honor, I feel it's important to

03:07:02 11   say something to clarify the record, if I could.

03:07:04 12           I believe that Mr. Mogin misspoke because the 29

03:07:08 13   categories are not by request.  The 29 categories are 29

03:07:14 14   subject categories that the plaintiffs have come up with that

03:07:18 15   they are asking the defendants to organize the documents

03:07:22 16   around.  They are not by request.  There are over 95 requests

03:07:26 17   here.

03:07:26 18           MR. MOGIN:  As you will see, your Honor, if you

03:07:28 19   look --

03:07:30 20           THE COURT:  All right.  We are using the term "best

03:07:34 21   practice."  Okay?  What I understand by best practice is the

03:07:38 22   plaintiff, you give -- under Rule 34, if you want documents in

03:07:46 23   a certain form of production, it's usually the plaintiff asks

03:07:50 24   for it in a certain way.  This wasn't a tricky question, this

03:07:54 25   wasn't into indexing.  Mr. Regard said, What I understand is

03:08:00  1  they did TIFF image, they did metadata.  I am saying, did the

03:08:06  2  plaintiff request that in the first place?

03:08:10  3          THE WITNESS:  I don't know --

03:08:12  4          THE COURT:  Hello?  We are not into indexing.  We are

03:08:14  5  not into 29 categories.  We are not into you can't read the

03:08:20  6  request to produce.

03:08:20  7          MR. MOGIN:  Your Honor, yes.  In the first request

03:08:24  8  for production of documents, a format was specified.

03:08:28  9          THE COURT:  Here?

03:08:28  10          MR. MOGIN:  Correct.  And the format was superseded

03:08:32  11  by the format order.  So the instructions in our request,

03:08:38  12  except for the indexing issue, is superceded by the ESI format

03:08:44  13  order entered by Judge Shadur as document --

03:08:46  14          THE COURT:  Which is document number?

03:08:48  15          MR. MOGIN:  244.

03:08:48  16          THE COURT:  244, thank you.

03:08:50  17          MR. NEUWIRTH:  And, your Honor, defendants have

03:08:52  18  already produced over a million pages of documents consistent

03:08:56  19  with that order.

03:08:56  20          THE COURT:  Good.  I didn't know that.  You know,

03:08:58  21  actually, Mr. Neuwirth, I didn't know that.  One million,

03:09:02  22  okay.  That helps.  Thank you.

03:09:06  23          MR. MAROVITZ:  So for the record --

03:09:08  24          THE COURT:  Now, the other indexing issue is

03:09:10  25  something we had said we are going to deal with, but not

| | | |
|---|---|---|
| 03:09:14 | 1 | during this hearing.  We weren't going to deal with it during |
| 03:09:18 | 2 | the hearing. |
| 03:09:20 | 3 | I think it's a different body of law.  They seem to |
| 03:09:24 | 4 | be saying that if I went with some computer-assisted, it would |
| 03:09:28 | 5 | be easier to do indexing with computer-assisted.  I don't know |
| 03:09:34 | 6 | if Mr. Regard knows anything about Boolean, if that helps |
| 03:09:38 | 7 | indexing.  I suppose that would be one question that somebody |
| 03:09:42 | 8 | could ask.  In 14 years, I have not had an indexing issue |
| 03:09:46 | 9 | before, so I wasn't as familiar with that area. |
| 03:09:50 | 10 | MR. MAROVITZ:  And, Judge, just to be clear, what I |
| 03:09:52 | 11 | hope to go over with Mr. Regard in terms of indexing is |
| 03:09:58 | 12 | Mr. Regard's opinion, as expressed before, that the quality of |
| 03:10:02 | 13 | the information that's been provided to the plaintiffs in this |
| 03:10:06 | 14 | case is consistent with the information that's kept within the |
| 03:10:12 | 15 | companies themselves. |
| 03:10:14 | 16 | So that's -- when we talk about indexing, that's -- |
| 03:10:18 | 17 | the encoding, that's where I am going.  I'm not intending to |
| 03:10:22 | 18 | tread upon the area that you just mentioned. |
| 03:10:28 | 19 | So what Mr. Regard is prepared to talk about is |
| 03:10:30 | 20 | whether or not the information that's provided in the |
| 03:10:32 | 21 | stipulation, which was agreed to by both parties, and the |
| 03:10:34 | 22 | order, Exhibits 13 and, as entered, 14, provide that sort of |
| 03:10:42 | 23 | robust, detailed information to the plaintiffs in this case. |
| 03:10:48 | 24 | And that's what he is prepared to testify about |
| 03:10:52 | 25 | today, not the 29 categories themselves, your Honor. |

| | | |
|---|---|---|
| 03:11:00 | 1 | THE COURT:  Okay. |
| 03:11:00 | 2 | BY MR. MAROVITZ: |
| 03:11:04 | 3 | Q.  So, in any event, just so it's clear, Exhibit -- if I |
| 03:11:08 | 4 | numbered them right on the set -- Mr. Regard, maybe you can |
| 03:11:10 | 5 | confirm this -- Exhibit 13 says document number 244 on the |
| 03:11:18 | 6 | top? |
| 03:11:18 | 7 | A.  Yes, sir. |
| 03:11:18 | 8 | Q.  And Exhibit 14 has the two lines with document 245 on the |
| 03:11:24 | 9 | top.  They're identical with an exception.  Exhibit 13, which |
| 03:11:30 | 10 | is document 244, actually contains the appendices that were |
| 03:11:36 | 11 | stipulated to and that appendix 1, as you can see, provides |
| 03:11:40 | 12 | the file formats and appendix 2 provides for ESI metadata and |
| 03:11:44 | 13 | coding fields.  And when Judge -- the original one I had |
| 03:11:56 | 14 | didn't have them both.  Let's focus on Exhibit 14. |
| 03:11:58 | 15 | A.  Yes, sir. |
| 03:11:58 | 16 | Q.  Okay.  Is this the ESI protocol that you understand was |
| 03:12:04 | 17 | agreed to by the parties and that's one of the issues that we |
| 03:12:08 | 18 | are discussing today? |
| 03:12:10 | 19 | A.  Yes, it is. |
| 03:12:12 | 20 | Q.  Under the agreed-to ESI protocol, do you understand that |
| 03:12:16 | 21 | ESI is being produced in a manner organized similarly to ESIs |
| 03:12:22 | 22 | kept in the ordinary course of business? |
| 03:12:24 | 23 | A.  My understanding of ESI that's been produced, this issue |
| 03:12:40 | 24 | of how the information is kept in the normal course of |
| 03:12:42 | 25 | business is a little misunderstood, I think, when we come to |

03:12:48  1   the connection of computer science and how people actually

03:12:50  2   interact with their systems.

03:12:54  3          The way information is physically kept by the

03:12:58  4   computer systems in the normal course of business is not only

03:13:04  5   very complicated; it's largely irrelevant to the user.  What

03:13:08  6   the user does is they go to the operating system or they go to

03:13:10  7   their software and they say, Show me all my emails listed by

03:13:16  8   sender, or, Show me all my emails listed by recipient.  They

03:13:20  9   don't go into the actual hard drive and say, Show me where the

03:13:24  10  ones and zeros are that represent the letters and words that

03:13:28  11  represent the pieces of my email.

03:13:32  12         So it's not the physical ones and zeros level the way

03:13:34  13  it was kept in the normal course of business, but it's

03:13:38  14  functionally the same as the way it was kept in the normal

03:13:42  15  course of business, and in some aspects, it's superior to the

03:13:44  16  way it was kept in the normal course of business.

03:13:48  17  Q.  And so let's go down a minute.  Take a look at appendix 2

03:13:54  18  to Exhibit 14.  It's the last three pages.  Which metadata

03:13:56  19  fields associated with the ESI in this case, to the extent

03:14:02  20  they exist on the documents, are defendants producing to the

03:14:06  21  plaintiffs?

03:14:06  22  A.  My understanding is the ones that are listed in this

03:14:10  23  appendix, appendix 2.

03:14:12  24  Q.  So let's start with, say, custodian metadata.  How could

03:14:16  25  custodian metadata be used by either party in the case to

03:14:20   1   organize the ESI?

03:14:20   2   A. Well, you've gone straight to custodian, which is actually

03:14:26   3   in a slightly different category than the rest of these

03:14:28   4   fields.

03:14:28   5   Q. If you'd like to start with Bates, that's fine.

03:14:32   6   A. Well, Bates is also slightly different than the ones in

03:14:34   7   this field as well. Let me explain the difference.

03:14:38   8        Your Honor, we have metadata that is already in our

03:14:40   9   computer system or in our application, so when I have an

03:14:44  10   email, behind the scenes, there are certain fields of

03:14:48  11   information that that email has that are listed here, such as

03:14:52  12   on the top of page 17, what is the message ID? That's a

03:14:56  13   number behind the scenes. That's a metadata that's in my

03:15:00  14   email system.

03:15:00  15        The first few fields in this list of appendix 2, the

03:15:04  16   Bates number, the beginning Bates, the ending Bates, the

03:15:08  17   beginning attachment, the ending attachment, the page count,

03:15:12  18   and the custodian, those are fields that are actually added

03:15:16  19   through the course of collection and the processing of

03:15:20  20   documents in a litigation context.

03:15:22  21        So they're metadata fields, but they are added after

03:15:26  22   they have collected. Those fields don't exist in the native

03:15:32  23   files in an organization or in a software system. They are

03:15:34  24   added.

03:15:34  25        So when you ask about custodian, that's a very useful

03:15:38 1  field, it allows you to organize documents by whom they were

03:15:42 2  collected from, but it's not something you would normally find

03:15:46 3  in an organization on the document when you go and collect it.

03:15:50 4  You have to add that.

03:15:50 5      So that's -- when I say in some ways this may be

03:15:54 6  superior to the way it's stored in an organization, if that's

03:15:58 7  a field that's valuable to you, you wouldn't find it in the

03:16:00 8  organizations where these documents were collected.

03:16:02 9  Q.  How about some of the fields that follow that?

03:16:06 10 A.  The rest of the fields, it's after the word "custodian,"

03:16:10 11 you start talking about size and path, those would be part of

03:16:14 12 the operating system.  The MD 5 hash is something that are the

03:16:20 13 fingerprint in this case, is actually something that's, again,

03:16:22 14 generated after the documents are collected as part of the ESI

03:16:26 15 process.  A useful field to identify documents that are

03:16:32 16 identical, even though they might have been held by multiple

03:16:34 17 custodians to facilitate efficiencies, but it's added after

03:16:38 18 the fact.

03:16:38 19     Native file link would point to the native file.  I

03:16:46 20 don't know that that's been used in this case.  There are

03:16:50 21 provisions in the protocol to produce native files in certain

03:16:54 22 circumstances.  I don't know if any have been produced yet,

03:16:56 23 but that would allow you to associate the TIFF image with the

03:17:00 24 underlying native file where that was appropriate.

03:17:02 25     And then we go down, we have some here that are

03:17:04   1   email-specific, so obviously non-emails would not have these

03:17:08   2   fields of the subject of the email, the date sent, the date

03:17:12   3   modified, although you can't have date modified for

03:17:18   4   non-emails.

03:17:18   5         Would you like me to continue to go down the list?

03:17:22   6   Q. No, unless there is something that's important for the

03:17:24   7   court to understand.

03:17:24   8         My question, unless there is something else,

03:17:26   9   Mr. Regard, is on this, is it possible for either side to go

03:17:30   10   ahead and sort documents by using these particular fields?

03:17:36   11   A. Well, I think that if you can organize documents and

03:17:42   12   retrieve them by these fields or by the content of the message

03:17:46   13   itself, whether it's in the native application or whether it's

03:17:48   14   in a subsequent production, you are afforded the same facility

03:17:52   15   and functionality that you would have in the native

03:17:56   16   application.

03:17:56   17   Q. Are plaintiffs or the defendants in this case able to

03:18:00   18   conduct key word searches on the ESI being produced by the

03:18:04   19   defendants pursuant to its protocol?

03:18:06   20   A. Well, depending on the specific software that they use,

03:18:12   21   there are many out there that do, the answer is absolutely.

03:18:14   22   Q. How would that be helpful in a case like this?

03:18:18   23   A. Well, there's been some discussion earlier. If you wanted

03:18:20   24   to find a particular phrase and you put that phrase in a

03:18:24   25   Boolean search, if could you identify the documents that had

03:18:26   1   that phrase, you could search within -- for communications

03:18:28   2   between two custodians by -- between two employees or two

03:18:34   3   recipients by searching for their names or their email

03:18:36   4   addresses. You could search for key words within the body or

03:18:40   5   the subject matter. You could rebuild conversations by

03:18:44   6   looking at the message ID, and you could link emails together

03:18:50   7   that were related to the same subject matter.

03:18:52   8   Q. And given those email messages which we discussed

03:18:56   9   oftentimes have attachments, are emails produced pursuant to

03:19:00   10   this protocol somehow associated with those attachments, to

03:19:04   11   the extent they exist?

03:19:04   12   A. Well, any documents that had a parent/child association

03:19:08   13   would normally be reflected in the Bates number range, the

03:19:14   14   beginning attachment, ending attachment range. In the systems

03:19:16   15   that we use or, at least the ones I am familiar with, for

03:19:20   16   managing litigation productions would then allow you to

03:19:24   17   reassociate the attachments with the parent document.

03:19:26   18   Q. In your opinion, Mr. Regard, is the ESI produced by

03:19:30   19   defendants at least as well organized as the ESI kept by

03:19:34   20   defendants in the ordinary course of their business?

03:19:36   21   A. It is at least as well organized -- well, I think it's

03:19:42   22   slightly better organized because, remember, this information

03:19:46   23   is being gathered from multiple locations within the

03:19:48   24   organization. We talked about custodial collections. If you

03:19:52   25   collected documents for a custodian from an email and from

03:19:56  1  their network and from their desktop, in this case, they'd all

03:20:00  2  be organized in one collection.

03:20:02  3  So the organization is at least as good, if not

03:20:06  4  better, and the functionality of retrieval is at least as

03:20:10  5  good, if not better.

03:20:12  6  Q.  Okay.

03:20:14  7  MR. MAROVITZ:  Judge, I had a question or two on the

03:20:18  8  29 different fields.  I will save that for another time, I

03:20:20  9  gather.

03:20:22  10  BY MR. MAROVITZ:

03:20:22  11  Q.  Finally, very quickly, Mr. Regard, backup tapes.  I want

03:20:30  12  to keep this at an extremely high level, because I understand

03:20:34  13  that part of today's process is not to delve into the

03:20:40  14  specifics of backup tapes, but I just want to ask you one

03:20:44  15  question or two questions as long as you're here.

03:20:46  16  What are the well accepted ESI best practices for the

03:20:50  17  consideration and treatment of backup tapes and other off-line

03:20:54  18  media?

03:20:54  19  A.  The best practice --

03:20:58  20  MR. MOGIN:  Objection, your Honor.  You know, Rule

03:21:00  21  34, Rule 26 are fairly specific about this, and I don't think

03:21:02  22  the best practices has anything whatsoever to do with

03:21:08  23  modifying those rules.

03:21:08  24  THE COURT:  This is another one that I am so confused

03:21:12  25  on as the fact-finder.  So this is -- I really don't -- I know

03:21:20   1   what's not reasonably -- I know what the law is.  I do not

03:21:24   2   know -- I could not -- because I could not plow through the

03:21:28   3   seven volumes, which you were kind enough to give me, of the

03:21:34   4   30(b)(6)s.  I wish I had a chart of the backup systems for

03:21:38   5   each of the -- if each of the people actually went to the

03:21:44   6   backup system with Nan Nolan, I don't know.  I don't know -- I

03:21:50   7   thought some of them changed systems halfway through.  I

03:21:56   8   thought some of the companies merged, they merged.

03:22:00   9        I mean, this, to me, is the mushiest part of the

03:22:04   10  whole case.  It is so -- and this goes to the search, as far

03:22:08   11  as I'm concerned.  This is not -- if you're using by

03:22:18   12  custodians, or you are using however many custodians you have

03:22:20   13  here, and you think these are the key custodians, I don't know

03:22:24   14  if the people who use that searched the backup tapes.  And I

03:22:28   15  don't know whether Mr. Regard is going to know that.

03:22:30   16       MR. MAROVITZ:  Right, and that's not where I intended

03:22:34   17  to go at all.

03:22:34   18       THE COURT:  Well, I think -- you know, I don't think

03:22:36   19  this -- I mean, to me, when I am looking at the search here, I

03:22:42   20  understand if it's disaster.  I mean, I don't mean to sound so

03:22:46   21  cocky.  But when we're talking about front level, do they have

03:22:52   22  all the documents that go to Nan Nolan, has somebody run them?

03:22:58   23  I thought somebody even ran searches on some of the backup

03:23:04   24  tapes.  I thought that's what Chris told me.  But I don't know

03:23:06   25  which of the six did that.

03:23:10  1    So, actually, this is an issue I really am sort of

03:23:12  2  concerned with, if Mr. Regard knows.

03:23:18  3    MR. MAROVITZ:  By all means.  I am interested --

03:23:22  4    THE COURT:  Do you know on Temple?  Let's do the one

03:23:24  5  which you started on.  What's Temple's story -- I don't mean

03:23:28  6  story, but, you know -- so Temple had 28/5 custodians.

03:23:44  7    THE WITNESS:  Yes, ma'am.

03:23:44  8    THE COURT:  Okay.  And you're calling it online,

03:23:46  9  off-line, somebody else is calling it active, non-active,

03:23:50  10  somebody else is calling it backup tapes.  Okay.  Just so we

03:23:54  11  know the same thing, your five custodians, were all of the

03:24:00  12  data for the years on the five custodians, was it on active

03:24:08  13  data?  Do you know?  And correct me if I am wrong in my

03:24:14  14  terminology here.

03:24:14  15    THE WITNESS:  Your Honor, I received the same

03:24:16  16  multiple volumes of 30(b)(6)s that you received.

03:24:20  17    THE COURT:  Well, you are getting paid to read.  I am

03:24:26  18  not getting paid for this.

03:24:28  19    THE WITNESS:  You know, I can't buy more hours,

03:24:30  20  though, so I still have the same limitation of time that your

03:24:32  21  Honor does.

03:24:34  22    I am familiar with some of the information that's

03:24:38  23  been provided.  I can't say I am familiar with all of it

03:24:42  24  sitting here today.

03:24:42  25    I know that Temple has considered the online data

| | | |
|---|---|---|
| 03:24:48 | 1 | that was available for its five custodians and has focused on |
| 03:24:50 | 2 | the online availability.  I don't know that anyone has done an |
| 03:24:56 | 3 | analysis of whether there are any gaps in that data and |
| 03:25:00 | 4 | whether or not any gaps, if they even did exist, would |
| 03:25:02 | 5 | potentially be filled in by any of the off-line systems. |
| 03:25:08 | 6 | MR. MAROVITZ:  And, Judge -- oh, go ahead.  I'm |
| 03:25:10 | 7 | sorry. |
| 03:25:10 | 8 | THE WITNESS:  And I think that's kind of where we |
| 03:25:12 | 9 | were going initially is that at this stage in collecting |
| 03:25:18 | 10 | documents and looking at documents, that is what I see |
| 03:25:20 | 11 | consistently in other cases as well, that people consider the |
| 03:25:24 | 12 | online systems first before -- in a knee-jerk reaction going |
| 03:25:32 | 13 | immediately to backup tapes and pulling tapes before they have |
| 03:25:36 | 14 | had a chance to assess what information we already had access |
| 03:25:38 | 15 | to, period. |
| 03:25:40 | 16 | MR. MAROVITZ:  Can I suggest this, your Honor?  This |
| 03:25:42 | 17 | is obviously an important issue for the court.  It makes sense |
| 03:25:44 | 18 | for us to get it right.  I don't want to speak for all the |
| 03:25:48 | 19 | other defendants here.  We'd be glad to submit something |
| 03:25:50 | 20 | shortly that would answer the question for each of the |
| 03:25:54 | 21 | defendants. |
| 03:25:58 | 22 | THE COURT:  I am just trying to -- I want Mr. Mogin |
| 03:26:02 | 23 | to have plenty of time here.  I am thinking about calling |
| 03:26:04 | 24 | Judge Holderman is what I am thinking about doing.  Has |
| 03:26:08 | 25 | anybody, based upon the schedule, any key custodian here, |

03:26:14  1  based upon the schedule, running off to a 5:00 o'clock plane?

03:26:18  2  If I go and change -- if I can get out of the 4:00 o'clock

03:26:22  3  meeting, does somebody have to run out the door?

03:26:26  4          MR. MAROVITZ:  No, we'd very much like to have the

03:26:28  5  opportunity to have Mr. Regard go home after this, if we can.

03:26:32  6          THE COURT:  I am not as worried about -- I want to

03:26:36  7  make sure Mr. Mogin has enough time now because under the

03:26:38  8  original schedule, he'd have 15 minutes, and he doesn't think

03:26:42  9  this is fair what is going on with Mr. Regard anyway.

03:26:46  10          MR. MAROVITZ:  Absolutely.

03:26:46  11          THE COURT:  I am going to go call the chief judge.

03:26:48  12  We will take five minutes.

03:26:52  13          MR. MOGIN:  Thank you, your Honor.

03:26:54  14     (Short break.)

03:32:56  15          THE COURT:  Okay.  I guess I have until about 4:30.

03:33:04  16  So what we will do -- you can all sit down.  Please sit down.

03:33:12  17  Thank you.

03:33:12  18          So I have until 4:30, so you are going to get --

03:33:14  19  whatever you need to do on cross you are going to be able to

03:33:18  20  do, and I will cut off my -- I will cut short my talk with the

03:33:22  21  lawyers.

03:33:24  22          MR. MOGIN:  Very good, your Honor.  Obviously, we are

03:33:24  23  not going to be able to get to Dr. Tenny.

03:33:30  24          THE COURT:  No, obviously we are not going to get to

03:33:34  25  Dr. Tenny today.

03:33:36 1    MR. MOGIN:  I do want to be clear that the defendants

03:33:38 2  took four hours to cross Dr. Lewis, so we would expect

03:33:42 3  something close to the time for Dr. Tenny.

03:33:44 4    THE COURT:  I -- you're right.  You're right.  You're

03:33:48 5  right, you're right.

03:33:48 6    MR. MAROVITZ:  Well, Judge, to be fair, we tried to

03:33:50 7  pick and choose that, as we did not cross-examine Mr. Hanners

03:33:52 8  at all.  We tried to use our time in the way we thought it

03:33:58 9  should be used.

03:33:58 10    THE COURT:  Right.  I didn't give people time to eat

03:34:00 11  either.  I mean, this whole darn thing wouldn't have happened

03:34:04 12  if this meeting was not here, or if we wanted to take a break,

03:34:08 13  I will go to the meeting and come back.  I am not -- that's

03:34:12 14  what I am trying to say.

03:34:14 15    I want to make sure that Mr. Mogin has -- he had a

03:34:18 16  lot of questions, it sounded like he had a lot of questions.

03:34:20 17  I am more than willing to stop, I will run downstairs and I

03:34:24 18  will come back up again.  And what I needed to say to you, I

03:34:28 19  could also say to you on a telephone.  It would be better

03:34:32 20  coming in person, but I could do it in a telephone conference

03:34:36 21  too.  Okay?

03:34:38 22    What we can't do is the testimony.  And Dr. Tenny,

03:34:44 23  you know, she should be given her own clean, fresh time if we

03:34:48 24  are going to need her.  Okay?

03:34:54 25    MR. MAROVITZ:  Thank you.

| | | |
|---|---|---|
| 03:34:54 | 1 | THE COURT:  So you finish up, and then we will have |
| 03:34:56 | 2 | Mr. Mogin, you know, do what he wants to do |
| 03:35:02 | 3 | cross-examine-wise. |
| 03:35:04 | 4 | MR. MAROVITZ:  Very good, your Honor. |
| 03:35:04 | 5 | BY MR. MAROVITZ: |
| 03:35:06 | 6 | Q.  In conclusion, Mr. Regard, as a matter of best practices, |
| 03:35:10 | 7 | who is in the best position to decide on an appropriate search |
| 03:35:14 | 8 | and retrieval protocol for the production of documents? |
| 03:35:18 | 9 | A.  As a matter of best practices and consistent with the work |
| 03:35:24 | 10 | that I do at the Sedona Conference -- |
| 03:35:26 | 11 | THE COURT:  Okay.  Number 6, Sedona principle |
| 03:35:30 | 12 | number 6, right? |
| 03:35:32 | 13 | THE WITNESS:  Yes, your Honor. |
| 03:35:32 | 14 | THE COURT:  Right. |
| 03:35:34 | 15 | THE WITNESS:  We are on the same page. |
| 03:35:34 | 16 | THE COURT:  Exactly.  So best practices, yes. |
| 03:35:40 | 17 | MR. MAROVITZ:  I could not have ended it better |
| 03:35:42 | 18 | myself, Judge. |
| 03:35:44 | 19 | We would like to formally move for admission for |
| 03:35:48 | 20 | Group Exhibit 12. |
| 03:35:48 | 21 | THE COURT:  I am thinking about that. |
| 03:35:50 | 22 | MR. MAROVITZ:  I understand that. |
| 03:35:50 | 23 | And also, just so the record is complete, I guess for |
| 03:35:54 | 24 | Defense Exhibit 14.  It's the court order.  I just want to |
| 03:35:58 | 25 | make sure it's in the record somewhere. |

03:36:00   1             THE COURT:  Do you have any objection to that?

03:36:02   2             MR. MOGIN:  None whatsoever.

03:36:02   3             THE COURT:  That's admitted, No. 14 is admitted.

03:36:04   4     (Above-mentioned exhibit was received in evidence.)

03:36:08   5             THE COURT:  Okay.  Thanks, Mr. Marovitz.

03:36:12   6             MR. MAROVITZ:  Thank you, your Honor.

03:36:12   7             MR. MOGIN:  I didn't get to complete my objection,

03:36:14   8     your Honor, but I think ultimately, you are the one who gets

03:36:18   9     to decide what the best practices with respect to search is.

03:36:20  10             THE COURT:  Right.

03:36:24  11                             - - -

03:36:24  12             DAN REGARD, CROSS-EXAMINATION

03:36:24  13     BY MR. MOGIN:

03:36:58  14     Q.  Mr. Regard.

03:37:02  15     A.  Good afternoon.

03:37:04  16     Q.  How are you?

03:37:04  17     A.  Fine.

03:37:06  18     Q.  So you are a lawyer, correct?

03:37:12  19     A.  I do not practice law, but I did go to law school, yes,

03:37:16  20     sir.

03:37:16  21     Q.  Are you admitted to practice law?

03:37:16  22     A.  I am.

03:37:18  23     Q.  Are you currently eligible to practice law?

03:37:20  24     A.  I am -- I don't know the answer to that.  I had to ask

03:37:26  25     Louisiana to put me on status so that I am not practicing so

03:37:32  1  that I don't have to continue to submit my CLE credits.  So I

03:37:36  2  can actively activate that if I chose to.

03:37:38  3  Q.  So are you currently on inactive status?

03:37:42  4  A.  I believe that's correct.

03:37:42  5  Q.  Let me show you what the Louisiana bar has to say about

03:37:48  6  that.  Do you mind?

03:37:50  7  A.  I hope they don't show that I am deficient on my CLE

03:37:54  8  credits.

03:37:54  9  Q.  In fact, they say you are ineligible to practice law.

03:37:58  10  Would that surprise you?

03:37:58  11  A.  No, sir.  I haven't been trying to practice law.

03:38:02  12  Q.  Well, they say you're -- you are not a member in good

03:38:06  13  standing, as they say, not an active member in good standing.

03:38:10  14  A.  Okay.

03:38:10  15        MR. MOGIN:  Your Honor, may I have this marked as

03:38:30  16  Plaintiff's 11.

03:38:40  17        THE COURT:  Thank you.

03:38:52  18        MR. MAROVITZ:  Judge, just one quick objection.  The

03:38:56  19  email that we received before this proceeding suggested that

03:39:00  20  all exhibits were to be tendered in advance by last Friday.

03:39:06  21  This exhibit, and maybe others, that Mr. Mogin intends to show

03:39:10  22  Mr. Regard were not among those.

03:39:12  23        THE COURT:  Mr. Mogin?

03:39:12  24        MR. MOGIN:  This is cross-examination, your Honor.

03:39:16  25        MR. MAROVITZ:  There was no exception for that in the

03:39:18  1  order we received, Judge.

03:39:20  2       MR. MOGIN:  In the normal course of proceedings, as

03:39:22  3  the court well knows, is that cross-examination --

03:39:24  4       THE COURT:  So you consider this impeachment?

03:39:26  5       MR. MOGIN:  I certainly do.

03:39:26  6       THE COURT:  Okay.

03:39:28  7  BY MR. MOGIN:

03:39:30  8  Q.  So, Mr. Regard, do you see the certificate from the state

03:39:36  9  bar of Louisiana?

03:39:36  10  A.  I do.

03:39:38  11  Q.  And it indicates that you have been ineligible to practice

03:39:42  12  since June of 2011, correct?

03:39:44  13  A.  That's what's reflected on here, yes.

03:39:48  14  Q.  And when did you first become aware that you were

03:39:50  15  ineligible to practice in Louisiana?

03:39:52  16  A.  Well, I haven't practiced in Louisiana, so it hasn't been

03:39:58  17  an issue for me to inquire on.

03:39:58  18  Q.  When did you first become aware that your status in

03:40:04  19  Louisiana was ineligible?

03:40:06  20  A.  I have actually been notified by Louisiana on several

03:40:10  21  occasions that my status was ineligible, and I have cured it

03:40:14  22  many times.  I haven't seen this until just now.

03:40:18  23  Q.  Are you suggesting that the certificate is inaccurate?

03:40:20  24  A.  No, sir.  I am trying to answer your question.

03:40:22  25  Q.  When did you first become aware, then, Mr. Regard, that

03:40:26    1   you were ineligible in Louisiana?

03:40:28    2   A.  Of this current ineligibility?

03:40:32    3   Q.  Have you been ineligible before?

03:40:34    4   A.  I just testified I have been noticed by Louisiana of

03:40:38    5   ineligibility in the past, and I have cured those.  The issue

03:40:42    6   arises from the status of active versus inactive and the

03:40:46    7   submission of CLE credits.

03:40:48    8   Q.  Well, this doesn't indicate inactive, it indicates

03:40:52    9   ineligible, does it not?

03:40:54   10   A.  This says ineligible on a -- I don't even know what the M

03:41:04   11   stands for, MCLE ineligible.

03:41:08   12   Q.  Correct.  You understand that Louisiana has a status of

03:41:10   13   active in good standing, ineligible to practice law for a

03:41:14   14   variety of reasons, and inactive, correct?

03:41:18   15   A.  I believe that's correct.

03:41:20   16   Q.  All right.  And your current status is ineligible,

03:41:24   17   correct?

03:41:24   18   A.  That's what this document says, yes, sir.

03:41:26   19   Q.  Okay.  Now, were you aware of that status at the time that

03:41:30   20   you submitted your resume and attached it to your report in

03:41:34   21   this matter?

03:41:34   22   A.  I was not.

03:41:36   23   Q.  Okay.  Did you make any attempt to verify that the

03:41:44   24   information that you had included in your resume with respect

03:41:46   25   to your status as a lawyer was correct?

03:41:50  1  A.  No, I relied upon my memory and my understanding of my

03:41:58  2  status at the time.

03:41:58  3  Q.  But you just testified, did you not, that Louisiana had

03:42:02  4  informed you several times that you were having problems

03:42:04  5  because of your non-compliance with their CLE requirements,

03:42:08  6  correct?

03:42:08  7  A.  And I also informed you that several times I have cured

03:42:12  8  that.  I was not aware of having ineligible status.  Thank you

03:42:16  9  for bringing it to my attention.  I will bring it up with the

03:42:20  10  bar this afternoon and cure it by tomorrow, if possible.

03:42:22  11  Q.  Okay.  Now, the fact that you are an attorney is an

03:42:26  12  important aspect of your consulting practice, is it not?

03:42:28  13  A.  The fact that I have a legal education I think is a

03:42:32  14  benefit to my clients and the work that I do, not whether I am

03:42:36  15  an active or inactive attorney.

03:42:38  16  Q.  But you do tell your clients that you are an attorney,

03:42:42  17  that you have a J.D., correct?

03:42:44  18  A.  I do have J.D.  I am an attorney.

03:42:46  19  Q.  And you tell them that you're licensed by Louisiana,

03:42:50  20  right?

03:42:50  21  A.  I do not tell them I am practicing law.  In fact, I make

03:42:54  22  it a point to tell them that I am not giving them legal

03:42:56  23  advice.

03:42:56  24  Q.  Well, it says, if we go to your website, does it not, that

03:43:00  25  you are admitted to the Louisiana state bar, just as it says

03:43:04    1    in your resume that you submitted here today?

03:43:08    2    A.   Yes, sir, just as this document says.

03:43:10    3    Q.   Well, is precision important, Mr. Regard?

03:43:14    4    A.   In what context, Mr. Mogin?

03:43:18    5    Q.   Do you mean to say that the word "precision" has more than

03:43:22    6    one meaning?

03:43:22    7    A.   In which context would you like to talk about precision?

03:43:26    8    Q.   Yes or no, can I have a yes-or-no answer to my question?

03:43:30    9    Does precision have more than one meaning?

03:43:32   10    A.   Precision can definitely have more than one meaning.

03:43:36   11    Q.   Is precision in one's testimony important when one is

03:43:40   12    acting as an expert witness?

03:43:44   13    A.   I don't associate precision with testimony.  I associate

03:43:48   14    accuracy with testimony.

03:43:48   15    Q.   Is there a difference between accuracy and precision?

03:43:52   16    A.   With respect to testimony, yes.

03:43:52   17    Q.   What's that difference?

03:43:54   18    A.   I don't know how to use the word precision in testimony.

03:43:56   19    "Accuracy" is in my opinion, trying to represent the

03:44:02   20    information as correctly as possible.

03:44:02   21    Q.   So precision could mean accuracy; is that right?

03:44:06   22    A.   I'm sorry?

03:44:06   23    Q.   Precision could mean accuracy?

03:44:08   24    A.   This is your definition.  If you want to put it into

03:44:12   25    context, I would be more than happy to talk about my

| | | |
|---|---|---|
| 03:44:14 | 1 | understanding. |
| 03:44:14 | 2 | Q. I put it in the context of your testimony. |
| 03:44:16 | 3 | A. Okay. |
| 03:44:16 | 4 | Q. Does precision mean accuracy? |
| 03:44:20 | 5 | A. Precision of my testimony as to what I am testifying |
| 03:44:26 | 6 | about? |
| 03:44:26 | 7 | Q. Yes. |
| 03:44:28 | 8 | A. Or precision as a word that I have used in the course of |
| 03:44:30 | 9 | my testimony? |
| 03:44:30 | 10 | Q. We will get to that. |
| 03:44:34 | 11 | A. Please repeat the question. |
| 03:44:36 | 12 | Q. The question is, is precision, as in accuracy, important |
| 03:44:40 | 13 | with respect to your testimony? |
| 03:44:42 | 14 | A. I believe that accuracy is important with respect to my |
| 03:44:46 | 15 | testimony. |
| 03:44:46 | 16 | Q. Is precision as in the context of accuracy important with |
| 03:44:50 | 17 | respect to your consulting practice? |
| 03:44:52 | 18 | A. Accuracy is important to my consulting practice, yes. |
| 03:44:58 | 19 | Q. Is precision of language important in both of those |
| 03:45:02 | 20 | contexts; that is, testifying as well as in your consulting |
| 03:45:08 | 21 | practice? |
| 03:45:08 | 22 | A. I think using the language correctly is important in |
| 03:45:10 | 23 | testimony and in consulting. |
| 03:45:12 | 24 | Q. Now, in e-discovery, and we will come back to what |
| 03:45:18 | 25 | e-discovery means, is it important that you're able to give |

03:45:20   1    your clients precise explanations of technology and their

03:45:28   2    options?

03:45:30   3    A.  I think you have to -- I think you have to judge what I

03:45:36   4    should give a client at a given time.  When you talk about

03:45:40   5    precision, precision can mean different things to different

03:45:44   6    clients in different contexts.  What is sufficiently precise

03:45:48   7    for one client may be insufficiently precise for another, so

03:45:48   8    you need to choose what level of detail you are going to

03:45:48   9    provide.

03:45:52  10           In terms of being accurate, I always strive to be

03:45:54  11    accurate.

03:45:56  12    Q.  Okay.  Isn't part of your job, as you see it as a

03:45:58  13    consultant, to explain in as clear language as you possibly

03:46:02  14    can to guide your clients through the e-discovery journey?

03:46:08  15    A.  Yes, sir.

03:46:08  16    Q.  So that they really do understand their options, correct?

03:46:12  17    A.  Hopefully.

03:46:14  18    Q.  Now, e-discovery, let's just talk about what "e-discovery"

03:46:18  19    means.  As you use e-discovery, what does that encompass?

03:46:24  20    A.  Electronic discovery deals with the handling of electronic

03:46:28  21    information in a litigation context.

03:46:32  22    Q.  Okay.  And in your experience, have you spent more time in

03:46:38  23    the preservation and collection end of e-discovery or in the

03:46:42  24    search technology?

03:46:44  25    A.  I don't know if I can separate those.  My cases tend --

| | | |
|---|---|---|
| 03:46:54 | 1 | the projects I work on tend to involve both of those, |
| 03:46:58 | 2 | preservation, collection, processing, searching, posting, |
| 03:47:02 | 3 | those all fall under the rubric of what we deal with on a |
| 03:47:08 | 4 | regular basis. |
| 03:47:10 | 5 | Q.  Well, let's go through your resume. |
| 03:47:12 | 6 | A.  Yes, sir. |
| 03:47:14 | 7 | Q.  Do you have your resume in front of you?  It's attached to |
| 03:47:18 | 8 | your report.  Is your report still there? |
| 03:47:18 | 9 | A.  I believe it's Exhibit A to my report. |
| 03:47:18 | 10 | Q.  Tell me, if you would, please, let's look at the bullet |
| 03:47:22 | 11 | points where you list your selected consulting experience. |
| 03:47:26 | 12 | A.  Yes, sir. |
| 03:47:26 | 13 | Q.  The first one, broker/dealer industry effectively reducing |
| 03:47:30 | 14 | the cost of e-discovery by 80 percent, et cetera, et cetera. |
| 03:47:36 | 15 | Were you asked in that particular instance to consult |
| 03:47:40 | 16 | on the search methodology or the search technology? |
| 03:47:44 | 17 | A.  I am trying to recall that particular engagement. |
| 03:48:00 | 18 | Yes. |
| 03:48:00 | 19 | Q.  Okay.  What did you do? |
| 03:48:02 | 20 | A.  We completely changed the historical archiving |
| 03:48:10 | 21 | architecture of that particular client; we developed a new |
| 03:48:14 | 22 | searching algorithm; we gave them the ability to retrieve |
| 03:48:16 | 23 | records on a custodial basis rather than a chronological |
| 03:48:20 | 24 | basis; and as is reflected here, we reduced their cost of |
| 03:48:26 | 25 | e-discovery. |

| | | |
|---|---|---|
| 03:48:26 | 1 | Q. So did you develop that searching algorithm? |
| 03:48:28 | 2 | A. Myself and my team, yes. |
| 03:48:32 | 3 | Q. What was your personal involvement, please. |
| 03:48:36 | 4 | A. Identification of the architecture, identification of the |
| 03:48:40 | 5 | replacement architecture, initial implementation of the |
| 03:48:46 | 6 | copying of historical data to the new architecture, working |
| 03:48:52 | 7 | with my programmers to help develop the algorithms and the |
| 03:48:56 | 8 | methodology to retrieve information, oversight of the project, |
| 03:49:02 | 9 | liaison with the client. |
| 03:49:04 | 10 | It's a while back, but that's what I recall. |
| 03:49:08 | 11 | Q. Working with the programmers, precisely what did you do? |
| 03:49:12 | 12 | A. Worked with them on the translation and transformation of |
| 03:49:22 | 13 | the emails, attachments, and documents from one format to |
| 03:49:26 | 14 | another; I worked with them on the programming of the SQL code |
| 03:49:36 | 15 | that was used to retrieve documents from the new environment; |
| 03:49:42 | 16 | gave them direction and worked with them to solve certain |
| 03:49:48 | 17 | programming problems. |
| 03:49:48 | 18 | Q. But the bottom line on that engagement, however, is that |
| 03:49:54 | 19 | with respect to your work with the programmers and your |
| 03:49:58 | 20 | development of the algorithm, it didn't have anything to do |
| 03:50:00 | 21 | with selection of responsive documents in this discovery |
| 03:50:04 | 22 | context, did it? |
| 03:50:06 | 23 | A. I wasn't tasked in that particular case to make a |
| 03:50:12 | 24 | relevancy determination, but a lot of the work that was being |
| 03:50:18 | 25 | done on that particular project was responding to SEC requests |

03:50:22  1  for information on a per-broker basis within specific trades,

03:50:26  2  within specific industries.  I don't know that there was a

03:50:32  3  relevancy determination that needed to be made.  Either

03:50:34  4  communications fit the criteria or they did not.

03:50:36  5          So we have the client make these extractions from

03:50:40  6  these historical archives based on the name of the broker, the

03:50:44  7  date ranges that were of interest, and the stocks that were of

03:50:46  8  interest.

03:50:48  9  Q.  So the answer is no?

03:50:48  10  A.  The answer is if you want to consider fulfilling the SEC's

03:50:54  11  requirements as a relevancy requirement, the answer would be

03:50:56  12  yes.  If you don't consider that a relevancy, the answer would

03:51:00  13  be no.

03:51:00  14  Q.  The question was having to do with the development of the

03:51:02  15  algorithm and the responsiveness aspect of development of the

03:51:06  16  algorithm.  And if I understood your testimony correctly, your

03:51:10  17  answer was that you didn't have anything to do with that

03:51:14  18  determination either because it wasn't necessary or because

03:51:16  19  you didn't do it; isn't that true?

03:51:20  20  A.  Again, I helped develop the algorithm, so maybe I am not

03:51:24  21  understanding your question.

03:51:24  22  Q.  Okay.  Was the purpose of the algorithm in any way to help

03:51:32  23  the parties determine responsiveness in the same sense as we

03:51:40  24  are discussing it here today?

03:51:40  25  A.  The algorithm was not to determine responsiveness.  The

03:51:44  1 algorithm was designed to retrieve records that the company

03:51:52  2 determined would be responsive to the request.  So the company

03:51:56  3 received the letter from the SEC, the company wanted to find

03:52:00  4 records that met these particular criteria as responsive.  The

03:52:04  5 algorithm executed that criteria against the data set.

03:52:08  6 Q.  What were the criteria?

03:52:08  7 A.  Name of the broker, name of the stock, date ranges.

03:52:14  8 Q.  Nothing else?

03:52:16  9 A.  Not that I recall.

03:52:16  10 Q.  Thank you.

03:52:18  11      Okay.  So the next one that you have listed here is,

03:52:26  12 In the pharmaceutical industry advised, client on

03:52:30  13 International Safe Harbor, discovery, and disclosure, correct?

03:52:34  14 A.  Yes, sir.

03:52:36  15 Q.  "International Safe Harbor," what does that refer to?

03:52:40  16 A.  That's a program that the United States has with Europe.

03:52:46  17 Europe has very stringent guidelines on the release of

03:52:52  18 personally identifiable information and sensitive information.

03:52:56  19 The EU has a data protection authority program that enforces

03:53:02  20 and interprets the data protection laws of Europe.  Europe has

03:53:08  21 designated that the United States does not sufficiently meet,

03:53:10  22 under our legal guidelines and paradigm, the privacy

03:53:16  23 requirements of Europe for personal information.

03:53:18  24      As an exception to that, Europe has designated that

03:53:22  25 if companies are listed and registered as safe harbor, that

03:53:28  1  they meet European guidelines and that those companies

03:53:32  2  individually, as opposed to the country of the United States,

03:53:36  3  are qualified to take possession of personal information

03:53:38  4  related to European citizens.

03:53:40  5  Q.  So you are giving legal advice on how to comply with an EU

03:53:48  6  policy?

03:53:48  7  A.  I spent a lot of time working with the Sedona Conference,

03:53:52  8  specifically working group 6, on developing international

03:53:58  9  programs and policies for the exchange of data.  I have worked

03:54:02  10  with the data protection authorities to get them to, if not

03:54:06  11  endorse, at least allow the publication of our recommended

03:54:10  12  best practices.  I don't know that I would call that legal

03:54:14  13  advice.  I am certainly helping clients understand what the

03:54:20  14  safe harbor options are in telling them what the data

03:54:24  15  protection authorities have told us with respect to those

03:54:26  16  programs.

03:54:26  17  Q.  In this particular engagement that you have listed here

03:54:30  18  under select consulting experience, the resume that you have

03:54:34  19  submitted to the court --

03:54:34  20  A.  Yes, sir.

03:54:34  21  Q.  -- did you provide legal advice on the International Safe

03:54:38  22  Harbor?

03:54:38  23  A.  No, sir.

03:54:40  24  Q.  So were you providing advice about best practices?

03:54:44  25  A.  Advice on the technologies used in that client to comply

03:54:52  1 | with their own discovery and disclosure obligations and the

03:54:56  2 | technology choices they made to enable a safe harbor program.

03:55:00  3 | I was working with in-house counsel, and I was

03:55:04  4 | working as a technology consultant to outside counsel.

03:55:06  5 | Q. Was Boolean search part of that?

03:55:08  6 | A. Not that I recall.

03:55:12  7 | Q. Was predictive coding?

03:55:16  8 | A. Not that I recall.

03:55:16  9 | Q. Were supervised learning?

03:55:18  10 | A. Not that I recall.

03:55:18  11 | Q. Statistically-ranked retrieval?

03:55:20  12 | A. Not that I recall.

03:55:22  13 | Q. Did you do any statistical analysis in connection with

03:55:24  14 | that particular engagement?

03:55:26  15 | A. Not that I recall.

03:55:28  16 | Q. Okay. So let's go down now to the tobacco industry

03:55:38  17 | engagement that you have listed here.

03:55:40  18 | It says that, In the tobacco industry, a team of 45

03:55:42  19 | team members to assess and advise client on legacy media and

03:55:48  20 | physical media inventories through collections, evaluations,

03:55:52  21 | remediation, and reintroduction into the normal document life

03:55:58  22 | cycle.

03:55:58  23 | Did I read that correctly?

03:56:00  24 | A. Yes, sir.

03:56:00  25 | Q. Then precisely what does reintroduction into the normal

03:56:04  1  document life cycle mean?

03:56:06  2  A.  That means enabling the client to take media, whether that

03:56:12  3  be tapes or disks, thumb drives, understand the information

03:56:16  4  that's on them, and for the client to create the process to

03:56:16  5  identify the information as to their normal document

03:56:28  6  rentention life cycle, which may mean to retain the

03:56:30  7  information, which may mean to expire the information.

03:56:32  8  Q.  Thumb drives, correct?  You mentioned thumb drives?

03:56:38  9  A.  Yes, sir.

03:56:38  10  Q.  What other type of non-desktop, non-server media was

03:56:44  11  involved in that particular engagement?

03:56:46  12  A.  Thumb drives, diskettes, external drives, mobile devices,

03:56:56  13  backup tapes.

03:56:58  14  Q.  And in connection with that engagement, you recognized,

03:57:00  15  did you not, the need to collect information from those

03:57:06  16  sources, true?

03:57:08  17  A.  This was not done in a legal context.  We collected the

03:57:14  18  physical devices.

03:57:16  19  Q.  This was a non-legal matter, had no connection to document

03:57:22  20  organizations, document retention policy?

03:57:24  21  A.  Eventually, this was assessed by the client of their own

03:57:30  22  document retention policies.  This was not associated with any

03:57:34  23  particular litigation.

03:57:34  24  Q.  So it's not an e-discovery matter?

03:57:38  25  A.  It's a matter that deals with electronic information.  The

03:57:44   1   client felt that it was related to their legal obligations for

03:57:46   2   document retention. I was engaged by counsel. It has -- you

03:57:52   3   can call it an ESI, an e-discovery issue or not, it's

03:57:56   4   electronically-stored information. It's the organization,

03:58:00   5   gathering, and identification of that information.

03:58:02   6   Q. Well, did you do anything in that particular engagement

03:58:06   7   with respect to Boolean search or supervised learning or

03:58:10   8   statistically-ranked retrieval?

03:58:14   9   A. I am considering whether or not we did any Boolean

03:58:22   10   searching on that.

03:58:24   11   I don't recall any Boolean searching. We definitely

03:58:26   12   did not do any machine run or statistical ranking.

03:58:30   13   Q. Did you perform any statistical analysis to verify your

03:58:34   14   findings or recommendations?

03:58:36   15   A. There was quantifications made of the data we were

03:58:42   16   finding, of the devices we were finding. There were

03:58:46   17   projections made of the expectations of data in other

03:58:48   18   locations. I don't recall if that was a statistical

03:58:54   19   evaluation, but statistics were generated and statistics were

03:58:58   20   reported.

03:58:58   21   Q. Okay. Let's try to speed this up a little bit.

03:59:02   22   Looking at all the engagements that you have listed

03:59:06   23   here on your resume, can you please identify any of them in

03:59:10   24   which you personally were involved in the development or

03:59:16   25   evaluation of either Boolean search methodologies like those

03:59:20  1  at issue here, supervised learning like those at issue here,

03:59:24  2  statistically-ranked retrieval like those at issue here, or

03:59:28  3  statistical evaluation?

03:59:30  4  A.  Well, first of all, do you want to go over the first case

03:59:36  5  we just talked about, because I think we just talked about

03:59:38  6  Boolean searching?

03:59:38  7  Q.  I am asking about the balance of them.  I don't want to go

03:59:42  8  back and replow ground.

03:59:46  9      MR. MAROVITZ:  Pardon me, your Honor.  Just a

03:59:48  10  clarifying objection.  Just so the record is clear, I don't

03:59:50  11  know whether Mr. Mogin is asking about, for instance, the

03:59:54  12  testifying experience as well or whether it's simply the first

03:59:58  13  consulting experience on pages 1 through 2.

04:00:02  14      MR. MOGIN:  I think the record is pretty clear that

04:00:04  15  it's select consulting experience, but if it isn't, I'll

04:00:08  16  clarify.

04:00:08  17      THE COURT:  Select, are you saying?

04:00:08  18      MR. MOGIN:  Select -- as Mr. Regard phrased it,

04:00:12  19  select consulting experience, pages 1 and 2 of his C.V., which

04:00:20  20  was tendered before as Defendants' 3.

04:00:24  21      THE WITNESS:  Your Honor, this is my select section

04:00:26  22  because it's not a complete list of all the projects I have

04:00:30  23  done over the last 20 years.

04:00:36  24      THE COURT:  I understand that.  Thank you.

04:00:44  25      THE WITNESS:  Was the question apply those

| | | |
|---|---|---|
| 04:00:58 | 1 | technologies or to develop those technologies? |
| 04:01:02 | 2 | BY MR. MOGIN: |
| 04:01:02 | 3 | Q.  Well, we will break it up. |
| 04:01:04 | 4 | A.  Okay. |
| 04:01:06 | 5 | Q.  You clearly want to tell us about both, so tell us first |
| 04:01:10 | 6 | about applied. |
| 04:01:10 | 7 | A.  In the majority of these cases, there were various forms |
| 04:01:14 | 8 | of document searching relying on Boolean technology.  I don't |
| 04:01:20 | 9 | know how long of detail you'd like me to go into those, but I |
| 04:01:24 | 10 | am more than happy to. |
| 04:01:24 | 11 | Q.  Did you develop the Boolean strings? |
| 04:01:32 | 12 | A.  Many times, we were given direction by counsel as to which |
| 04:01:36 | 13 | strings to develop.  We executed the strings.  Counsel would |
| 04:01:40 | 14 | tell us to look for key words or key concepts, and we would |
| 04:01:42 | 15 | develop the strings and execute those. |
| 04:01:44 | 16 | Q.  I want to get back just for a moment to precision and |
| 04:01:46 | 17 | accuracy.  You used the term "we," and my question was you. |
| 04:01:54 | 18 | A.  I often lead people on projects that I manage. |
| 04:01:58 | 19 | Q.  Did you on any of these engagements develop the Boolean |
| 04:02:02 | 20 | strings? |
| 04:02:02 | 21 | A.  I think I just said it's usually counsel that indicates |
| 04:02:08 | 22 | which words or which terms they want us to search for, they |
| 04:02:12 | 23 | want me to search for, and I will either execute it myself or |
| 04:02:16 | 24 | I will have my team execute it. |
| 04:02:16 | 25 | Q.  So they ask you to operate the tool and execute a |

04:02:20  1  preconstructed Boolean search?  Is that what you are saying?

04:02:24  2  A.  No, they ask me which words to organize, and I build that

04:02:28  3  into a Boolean language and then execute it against the

04:02:32  4  project.

04:02:32  5  Q.  They ask you which words to organize?

04:02:36  6  A.  They provide me the words, and I organize them in a

04:02:38  7  Boolean string.

04:02:40  8  Q.  So does that mean that you put in the operators?

04:02:44  9  A.  Sometimes, or the parentheses.

04:02:46  10  Q.  What else?

04:02:48  11  A.  Sometimes that's all I need to do.

04:02:50  12  Q.  Okay.  And how about with respect to -- so that's both

04:02:58  13  your testimony with respect to both development and

04:03:00  14  application, or is there more to that?

04:03:02  15  A.  Let me finish looking at this, please.

04:03:30  16       On page 3, there are some projects that I had more

04:03:36  17  significant hands-on programming and development with.

04:03:38  18  Q.  And which are those, please.

04:03:40  19  A.  In the first one, design, implemented a bank transaction

04:03:46  20  analysis and tracing database.  That involved the conversion

04:03:50  21  of paper documents that had been scanned in, the (inaudible)

04:03:56  22  of those documents, the building of a repository, and the

04:04:00  23  development of metadata analysis to correlate specific types

04:04:02  24  of transactions related to that investigation.

04:04:06  25  Q.  Please explain a little bit more about the development of

| | | |
|---|---|---|
| 04:04:10 | 1 | metadata analysis. |
| 04:04:10 | 2 | A.  The reverse -- for lack of a better term, the reverse |
| 04:04:18 | 3 | engineering of specific documents that were in paper, scanning |
| 04:04:22 | 4 | them in, these are trade confirmations in the broker/dealer |
| 04:04:26 | 5 | industry, converting that information back into tabular |
| 04:04:32 | 6 | format, and developing analyses to correlate dates, dollars, |
| 04:04:38 | 7 | stocks, and brokers, and clients to investigate specific |
| 04:04:42 | 8 | transactions. |
| 04:04:42 | 9 | Q.  Do you mean to say that you took paper documents, scanned |
| 04:04:46 | 10 | them in, and then engaged in further activity with respect to |
| 04:04:52 | 11 | that? |
| 04:04:52 | 12 | A.  Yes. |
| 04:04:52 | 13 | Q.  So the material that you were originally working with was |
| 04:04:58 | 14 | not ESI, correct? |
| 04:04:58 | 15 | A.  It became electronically stored once I converted it to |
| 04:05:04 | 16 | electronic information. |
| 04:05:04 | 17 | Q.  When did this take place? |
| 04:05:06 | 18 | A.  This would have been about 2001. |
| 04:05:10 | 19 | Q.  So in 2001, this bank didn't have ESI, didn't have the |
| 04:05:18 | 20 | confirmations in electronically-stored information? |
| 04:05:20 | 21 | A.  They weren't electronic information records for the |
| 04:05:26 | 22 | information I was dealing with at the time.  I don't find it |
| 04:05:28 | 23 | unusual at all to find paper documents in organizations and |
| 04:05:32 | 24 | not have the electronic sources of that, especially back in |
| 04:05:36 | 25 | 2001. |

04:05:36 1 Q. Bank transactions in 2001, you don't find that unusual, in

04:05:42 2 your experience?

04:05:42 3 A. In that particular case, the records were not available

04:05:46 4 electronically, so we dealt with what we had.

04:05:48 5 Q. Okay. Now, did you develop Boolean search strings in that

04:05:54 6 case?

04:05:54 7 A. Boolean search strings and much more complicated

04:05:58 8 algorithms as well. There was significant programming

04:06:02 9 involved in that case. I did it myself.

04:06:04 10 Q. Okay. What programming language did you use?

04:06:08 11 A. The majority of this was done in Visual Basic For

04:06:14 12 Applications, VBA.

04:06:16 13 Q. Now, is VBA a language that you learned when you got your

04:06:20 14 BS in computer science?

04:06:22 15 A. No, sir, I learned it subsequent.

04:06:24 16 Q. Are there other computer languages that are currently in

04:06:28 17 use that you are conversant -- strike that -- that you are

04:06:34 18 able to program?

04:06:36 19 A. I don't do a lot of programming today, but the programming

04:06:40 20 that I do is mostly in SQL.

04:06:42 21 Q. Okay. So are you familiar with C++?

04:06:44 22 A. I am.

04:06:46 23 Q. Can you program in C++?

04:06:48 24 A. I have in the past. I probably could again if I needed

04:06:52 25 to.

04:06:52  1  Q. When was the last time you did it?

04:06:54  2  A. Probably in the late '90s, early 2000s.

04:07:00  3  Q. How about Java, have you able to program in Java?

04:07:04  4  A. Never had a need to program in Java.

04:07:06  5  Q. Do you know the source programs, the source language

04:07:10  6  that's used in constructing any of the tools that were used by

04:07:14  7  defendants that are represented on your chart in Exhibit 4?

04:07:18  8  A. No, sir.

04:07:18  9  Q. Do you think that reviewing the algorithms was in any way

04:07:24  10  important to your analysis today?

04:07:24  11  A. No, sir.

04:07:26  12  Q. Did you make any attempt to look at the algorithm, to look

04:07:30  13  at the source code?

04:07:30  14  A. Since I didn't feel that I needed to, I did not make an

04:07:34  15  attempt to do so.

04:07:34  16  Q. All right. Now, on this page 3, were there other --

04:07:40  17  strike that.

04:07:40  18      Let's go back to this bank transaction engagement.

04:07:44  19  A. Yes, sir.

04:07:44  20  Q. Did you engage in anything having to do with

04:07:48  21  statistically-ranked retrieval?

04:07:56  22  A. In that particular case, no.

04:08:00  23  Q. Okay. How about use of supervised learning technologies?

04:08:06  24  A. Also not in that case.

04:08:08  25  Q. So are there -- there's, what, three more cases here that

04:08:12  1  are listed on page 3.  Are any of those relevant to what we

04:08:16  2  have been talking about?

04:08:16  3  A.  These are in the back of my resume.  I am still trying to

04:08:56  4  recall exactly which clients --

04:08:58  5  Q.  Take your time.

04:09:08  6  A.  In the second to last one.

04:09:10  7  Q.  That would be the (inaudible) the largest map of

04:09:12  8  restatement in U.S. history?

04:09:14  9  A.  That would be the one.

04:09:18  10        I was involved in a lot of hands-on programming and

04:09:22  11  development of the financial system analysis that we did in

04:09:26  12  that case.  It would not be what I would consider Boolean

04:09:30  13  searching because this was largely financial and mathematical

04:09:34  14  transactions, but we ended up -- and I say "we," myself and my

04:09:38  15  team -- ended up, as it states here, forming over 8,000

04:09:42  16  different analyses and reports off of that data.

04:09:46  17        So to the extent you're programming and data

04:09:50  18  management, that was a very significant effort over many, many

04:09:54  19  months.

04:09:56  20        In the last matter --

04:09:58  21  Q.  Hold on a second.  When did that particular engagement

04:10:04  22  take place?

04:10:04  23  A.  That engagement took place in the year 2000.

04:10:06  24  Q.  And this says, Delivered over 8,000 separate data sets for

04:10:12  25  further analysis; is that correct?

04:10:12    1   A.   It is.

04:10:14    2   Q.   So did you deliver it to some other person or organization

04:10:18    3   for further analysis?

04:10:18    4   A.   There was a forensic analysis team of about a hundred

04:10:24    5   people there.  We were delivering analyses to them.

04:10:30    6   Q.   Is that the team that you were leading?

04:10:32    7   A.   No, sir.  I was leading a team of 30, as it says here, a

04:10:36    8   technology team supporting a forensic accounting team.

04:10:40    9   Q.   So your primary job there was to lead the team that was

04:10:42   10   engaged in providing technical support, correct?  And the

04:10:46   11   technical support consisted mostly of collection work, right?

04:10:50   12   A.   There was a lot of collection work here, but I think you

04:10:56   13   understate the role that I played on that team.  This entire

04:11:00   14   effort was an effort that I designed:  The coalescing of

04:11:04   15   information from over 18 different subledgers and general

04:11:08   16   ledger systems within this company, the inspection, the

04:11:14   17   categorization, the identification of backup tapes to restore

04:11:20   18   the analysis of online data that was available, the analysis

04:11:22   19   of off-line data, the restoration of that data.  This was a

04:11:26   20   project that I not only led, I designed and had a lot of

04:11:32   21   hands-on work with.

04:11:32   22   Q.   This is another engagement where what you just called

04:11:36   23   off-line data proved to be important, I take it?

04:11:38   24   A.   In this case, it was, yes.

04:11:40   25   Q.   And what were some of the criteria that you used to make a

04:11:42　1　determination then in that case off-line data might be

04:11:46　2　important?

04:11:46　3　A.　The company was in the process of making a restatement

04:11:54　4　over five years of financial information.　Their financial

04:11:58　5　system had kept its annual information on their tape library

04:12:06　6　system that they used at the time, and so we went there to get

04:12:10　7　the information in order to do the analysis.

04:12:14　8　Q.　So because there was a five-year period at issue, that was

04:12:16　9　the primary driver between looking at what was off-line; is

04:12:16　10　that correct?

04:12:20　11　A.　No, it was a correlation in that particular case of the

04:12:22　12　online availability of information, the off-line availability

04:12:24　13　of information, and the cost and effort of restoring that

04:12:30　14　information.

04:12:32　15　Q.　Well, could you have responded over that five-year period

04:12:36　16　without going to the off-line information?

04:12:36　17　A.　Yes.

04:12:38　18　Q.　But you chose not to.　Why not?

04:12:40　19　A.　The company chose not to.　They were given options, and

04:12:44　20　they chose to restore.

04:12:44　21　Q.　They instructed you in that regard?

04:12:48　22　A.　I'm sorry?

04:12:48　23　Q.　In other words, you didn't have the discretion; is that

04:12:48　24　right?

04:12:52　25　A.　I didn't make the choice to restore the off-line data.

04:12:54  1  That was made at a higher pay scale than I was at.

04:12:58  2  Q.  Did you advise them to go back and go after the off-line

04:13:04  3  data?

04:13:04  4  A.  No.  I advised them of their options.

04:13:06  5  Q.  Did you advise them that it was your opinion that they

04:13:12  6  should not go and look at the off-line data?

04:13:14  7  A.  I did not.  I merely gave them their options.

04:13:18  8  Q.  Okay.  So the five-year time period was just -- in your

04:13:22  9  analysis, that wasn't important for the determination of

04:13:26  10  whether to look at the off-line data; is that right?

04:13:30  11  A.  That was the time frame that I was instructed to work on.

04:13:38  12  That was the time frame that was relevant to the audit and the

04:13:40  13  restatement.

04:13:42  14  Q.  Okay.

04:13:42  15  A.  That's the time frame of information that either was on

04:13:46  16  the online systems or was augmented from data from the

04:13:48  17  off-line systems.  There were other choices.  That's what the

04:13:52  18  client decided to do.

04:13:52  19  Q.  Have you ever heard the expression, preserve broadly and

04:13:58  20  produce narrowly?

04:13:58  21  A.  I have.

04:14:00  22  Q.  What does that mean to you?

04:14:02  23  A.  That means that many times when companies are in a

04:14:10  24  litigation, they make the decision to preserve information

04:14:14  25  today, and because information is often commingled between

04:14:20  1  relevant and non-relevant information, that preservation can

04:14:24  2  include non-relevant information and relevant information.  So

04:14:28  3  if you are preserving, you end up collecting -- you end up

04:14:32  4  locking down and ultimately collecting more information than

04:14:34  5  you ultimately intend to produce.

04:14:36  6  Q.  And what does the "produce narrowly" part of that mean?

04:14:42  7  A.  That means to try to produce the amount of information

04:14:46  8  that's actually relevant to the case and not to overproduce.

04:14:48  9  Q.  Okay.  And does that entail taking the restrictive view of

04:14:54  10  what's relevant to the case?

04:14:54  11  A.  It's not a review that I take either way.  It's to try to

04:15:02  12  tailor the ultimate handling of production of documents to the

04:15:06  13  needs of the case.

04:15:06  14  Q.  Do you advise clients to preserve broadly and produce

04:15:10  15  narrowly?

04:15:10  16  A.  I have used that phrase in the past.

04:15:14  17  Q.  Is that how you advised the clients here?

04:15:16  18  A.  I didn't advise the clients here on preservation or

04:15:20  19  production.

04:15:20  20  Q.  Okay.  In other words, you were just talking in the

04:15:24  21  beginning about collection; is that right?

04:15:28  22  A.  I'm sorry?

04:15:28  23  Q.  You didn't tell -- you didn't talk to the clients here

04:15:32  24  about preservation; is that correct?

04:15:32  25  A.  I didn't talk to them about the scope of preservation, I

04:15:36 1 didn't talk to them about the scope of collection. They asked

04:15:38 2 me for advice on techniques, on how to execute a quality

04:15:42 3 selection, and I gave them advice.

04:15:44 4 Q. Did you give advice with respect to other aspects of

04:15:48 5 preservation?

04:15:48 6 A. I did not.

04:15:48 7 Q. So you gave limited advice with respect to certain aspects

04:15:52 8 of collection; is that right?

04:15:54 9 A. That's a fair statement, yes.

04:15:56 10 Q. Okay. So when you tell somebody in the context of your

04:16:00 11 job as an e-discovery consultant who advises those clients,

04:16:08 12 that you're also an attorney, that they should preserve

04:16:14 13 broadly but produce narrowly, do you think that that has any

04:16:20 14 impact on the decisions that they make as to relevance and

04:16:26 15 what should be produced because it's responsive or not?

04:16:30 16 A. I certainly hope not.

04:16:36 17 Q. Do you think it does?

04:16:36 18 A. It's not the intent of what I do. I can't tell you how it

04:16:40 19 lands. I can only tell you the advice that I give.

04:16:44 20 Q. Well, you say you can't tell me how it lands, but

04:16:46 21 certainly you've heard clients' responses when you make that

04:16:50 22 statement, haven't you?

04:16:50 23 A. In what manner, Mr. Mogin?

04:16:56 24 Q. Well, such as, you know, That's a good idea, producing

04:16:58 25 narrowly?

04:17:00  1  A.  If they asked me to explain what producing narrowly means,

04:17:04  2  I explain the way that I just said.  It means to produce the

04:17:06  3  information in a way that's tailored to meet your obligations

04:17:10  4  and not to overproduce.

04:17:12  5  Q.  Isn't it true, sir, that when you have advised people to

04:17:14  6  produce narrowly, that you know that they have taken that to

04:17:18  7  mean to be restrictive in their interpretation of

04:17:22  8  responsiveness?

04:17:24  9  A.  Not at all.

04:17:24  10  Q.  Never?

04:17:24  11  A.  Not that I'm aware of.

04:17:26  12  Q.  Now, let's go, if we can, please, to your testifying

04:17:34  13  experience.

04:17:36  14  A.  Yes, sir.

04:17:36  15  Q.  Which I believe is page 4 of the exhibit; that is,

04:17:40  16  Defendants' 3.

04:17:42  17  A.  Yes, sir.

04:17:42  18  Q.  It's a lengthy list.  Can you please tell me if there are

04:17:46  19  any engagements on that list where you testified -- that is,

04:17:52  20  like you're testifying today, not by declaration or affidavit,

04:17:56  21  live testimony -- about any of the subjects we are talking

04:18:02  22  about; that is, Boolean search, supervised learning,

04:18:06  23  statistically-ranked retrieval, statistical validation?

04:18:10  24  A.  Most of my testimony is not in live court, most of my

04:18:18  25  testimony is by deposition, but let me look over this list.

04:18:36 1        You only want court testimony, or would live

04:18:38 2 deposition suffice for your needs?

04:18:40 3 Q.  We are going to break it up into categories, so let's

04:18:44 4 start with court testimony.

04:18:46 5 A.  Okay.  In the In Re Devinney (phonetic) Universal matter,

04:19:12 6 which is about two-thirds of the way down the list.

04:19:14 7 Q.  Just a second.  Let me locate it.

04:19:18 8        Okay.  Thank you.

04:19:18 9 A.  I testified as to the construction of SQL query language

04:19:24 10 and the results that the defendants had produced in that

04:19:26 11 particular case with respect to database tables and methods

04:19:32 12 that they used to produce those tables.

04:19:32 13 Q.  Now, database tables are what's referred to as structured

04:19:36 14 data, correct?

04:19:38 15 A.  Most data is structured.  Databases are structured data.

04:19:42 16 In this case, they contained unstructured data within the data

04:19:48 17 fields.

04:19:48 18 Q.  Is email considered structured data or unstructured data?

04:19:52 19 A.  That's a great question.  I think of email as a database

04:19:56 20 that contains unstructured information.

04:19:58 21 Q.  Is that the consensus view in the computer forensics

04:20:04 22 industry?

04:20:04 23 A.  I couldn't tell you.  I haven't had it polled.

04:20:06 24 Q.  Well, how about this.  You know a lot about best

04:20:10 25 practices.  Is it best practices to treat email as a database?

04:20:12    1    A.  Most people treat email in the same fashion they treat

04:20:18    2    loose files, which is as an unstructured container.

04:20:22    3    Q.  And you mentioned seraquel (sic).  Would you explain to

04:20:26    4    the court what that is?

04:20:26    5    A.  I didn't catch that word.  Could you repeat the question?

04:20:28    6    Q.  Sequarrel, seraquel, I can't pronounce it.  You mentioned

04:20:32    7    that it's --

04:20:32    8    A.  "Sequel."

04:20:34    9    Q.  Go ahead.  If you could explain to the court what SQL is.

04:20:38    10   A.  "Sequel" is a pronunciation of the acronym SQL, Structured

04:20:44    11   (sic) Query Language.  It's a programming language methodology

04:20:46    12   used to handle electronic database systems.

04:20:50    13   Q.  And so did you use Boolean queries?

04:20:52    14   A.  SQL has many of the same aspects of Boolean because you

04:21:02    15   are using the same operations.  So when I say SQL, that

04:21:06    16   programming language often uses the same structure -- or the

04:21:10    17   same equivalent structure in terms that you can find in

04:21:14    18   Boolean searching.

04:21:16    19   Q.  Did you use Boolean key word searching in connection with

04:21:22    20   that engagement?

04:21:22    21   A.  Yes, sir.

04:21:22    22   Q.  And did you develop the Boolean search strings?

04:21:26    23   A.  I did.

04:21:26    24   Q.  And how did you go about doing that?

04:21:28    25   A.  By having discussions with counsel, reading the pleadings,

04:21:40   1   developing a plan with counsel as to how to derive the

04:21:44   2   information counsel is interested in either having it analyzed

04:21:48   3   or looking at, developing the strings against the database

04:21:50   4   fields to extract the information and prepare it for counsel.

04:21:54   5   Q. What kind of information was in the database?

04:21:54   6   A. My recollection is largely financial transactions.

04:22:02   7   Q. So what sort of key words were you using in order to

04:22:06   8   examine financial transactions in the database?

04:22:08   9   A. The name of the financial account, so if you had a chart

04:22:14   10   account, the name of that account, or maybe the account

04:22:16   11   number; the extent that there was a payee or a creditor, the

04:22:20   12   name of that client or that creditor; to the extent there were

04:22:24   13   descriptive texts within the fields as to what transactions

04:22:28   14   took place --

04:22:28   15   Q. I'm sorry. As to?

04:22:30   16   A. If there was a descriptive term within the description

04:22:34   17   field for that transaction, as to why that transaction took

04:22:38   18   place.

04:22:38   19   Q. Thank you.

04:22:38   20   A. We'd be looking for text. If there was commentary from

04:22:42   21   the accounting department as to the backup meaning behind a

04:22:50   22   particular transaction.

04:22:50   23   So, again, there was dated information, there was

04:22:54   24   currency information, there was free-flowing, unstructured

04:22:58   25   text information both with the transaction nature and the

04:23:00  1  commentary.

04:23:02  2  Q. Were you looking for coded phrases?

04:23:04  3  A. A lot of the information in financial database use

04:23:10  4  acronyms and codes.

04:23:12  5  Q. Codes as in the sense of someone trying to obscure or

04:23:16  6  hide?

04:23:18  7  A. Yes, sir.

04:23:20  8  Q. And how did you go about formulating Boolean queries for

04:23:26  9  that particular information?

04:23:26  10  A. We started by looking for information with respect to

04:23:32  11  particular financial accounts, particular participants in

04:23:34  12  those transactions. We produced those to counsel. We

04:23:40  13  identified particular transactions of interest to them and

04:23:44  14  asked us to see if we could find more transactions of that

04:23:46  15  nature. We would look at the terminology that was used to try

04:23:48  16  to find additional transactions using the same terminology.

04:23:54  17  Q. Are you saying that there were certain transactions that

04:23:56  18  were material to the investigation that might have been hidden

04:23:58  19  in some way?

04:24:00  20  A. That might have -- excuse me?

04:24:02  21  Q. That somebody might have tried to hide in some way and

04:24:06  22  that's what you were searching for?

04:24:06  23  A. Not the transactions were hidden, but that the

04:24:08  24  transactions were in the accounting system, and we were trying

04:24:10  25  to narrow down the accounting system to just those

04:24:14   1   transactions.

04:24:14   2   Q. So you were using the Boolean search in order to narrow;

04:24:14   3   is that correct?

04:24:20   4   A. It is.

04:24:22   5   Q. And in that particular regard, did you use

04:24:30   6   statistically-ranked retrieval?

04:24:32   7   A. Not that I recall.

04:24:32   8   Q. Did you use supervised learning?

04:24:34   9   A. Not in that matter, no.

04:24:38   10   Q. Did you do a statistical evaluation of the validity of

04:24:42   11   your own searches?

04:24:42   12   A. No, sir.

04:24:46   13   Q. Did you engage in random sampling?

04:24:48   14   A. No, sir.

04:24:52   15   Q. Let's go back then. Any other court testimony that's

04:24:58   16   pertinent?

04:25:00   17   A. I don't see any other testimony on Boolean search

04:25:38   18   terminology or machine language.

04:25:40   19   Q. How about where you were engaged in validation,

04:25:46   20   statistical validation searches?

04:25:50   21   A. No, sir.

04:25:50   22   Q. How about -- then let's talk about the testimony that you

04:25:56   23   have given, the deposition testimony that you have given that

04:25:58   24   might cover those sections. Can you identify any such

04:26:02   25   engagements on this exhibit?

04:26:02 1 A. There was significant testimony in deposition in the

04:26:14 2 Gentex v. Armor Holdings matter, which is the second one from

04:26:18 3 the top. That involved a significant amount of searching.

04:26:22 4 Q. Okay. And you were the forensics expert and computer

04:26:30 5 science expert in that particular case?

04:26:32 6 A. Yes, sir.

04:26:32 7 Q. What was your role as the computer forensics expert?

04:26:36 8 A. My role in computer forensics was to assess, if possible,

04:26:40 9 the quantity of data that had been taken from company A by

04:26:46 10 exiting employees in transition to company B. It was also to

04:26:52 11 identify the amount of information in company A's systems that

04:26:56 12 had been maliciously deleted by the exiting employees. It was

04:27:02 13 to identify the manner in which the exiting employees

04:27:04 14 maintained that data after they left the company and how they

04:27:08 15 shared it with each other, with other people within company B.

04:27:12 16 It was to examine how they had deleted information in company

04:27:16 17 B after they had been filed, after they arrived and used the

04:27:24 18 information.

04:27:24 19 It was to examine how computer forensic examiners who

04:27:28 20 were hired by company B conducted their investigation. It was

04:27:32 21 to evaluate the quality of the results of the forensic

04:27:36 22 examiners hired by company B in their investigation. It was

04:27:40 23 to examine a purge activity conducted by the forensic

04:27:46 24 examiners at company B and the quality of that purge. It was

04:27:48 25 to render an opinion on the sustained presence of additional

04:27:52   1   information from company A still in the possession of

04:27:56   2   company B.

04:27:56   3   Q.  So if I understand correctly, what you're saying is that

04:28:00   4   -- were you representing -- you were testifying on behalf of

04:28:04   5   defense in that case?

04:28:04   6   A.  No, sir, plaintiff.

04:28:06   7   Q.  So what you were doing on behalf of the plaintiff then was

04:28:10   8   you reviewed the work of the defense investigators and

04:28:14   9   testified about certain aspects of that?

04:28:16   10  A.  In part, yes.

04:28:16   11  Q.  And what did you do as the computer science expert that's

04:28:20   12  different than what you just described as the computer

04:28:24   13  forensics expert?

04:28:24   14  A.  One of the of the issues in this case was the accusation

04:28:28   15  or the theft of spreadsheets and databases from company A that

04:28:36   16  were then used at company B.  And from a computer science

04:28:40   17  perspective, my testimony included the identification of this

04:28:44   18  information as intellectual property or the kind of

04:28:48   19  information that could be identified as intellectual property;

04:28:52   20  and also to trace the evolution of these documents as they

04:28:56   21  were modified at company B and ultimately as they were used to

04:29:00   22  create fresh instances of databases and spreadsheets that were

04:29:04   23  in continuous use at company B.

04:29:06   24  Q.  So, precisely, did you create any Boolean search strings?

04:29:10   25  A.  Yes, sir.

| | | |
|---|---|---|
| 04:29:12 | 1 | Q. Okay. Tell me what you did as far as creating Boolean |
| 04:29:18 | 2 | search strings. How many? |
| 04:29:20 | 3 | A. Hundreds. |
| 04:29:22 | 4 | Q. Hundreds of different strings? |
| 04:29:24 | 5 | A. Yes, sir. |
| 04:29:24 | 6 | Q. You're not referring to hundreds of stems, you're talking |
| 04:29:28 | 7 | about hundreds of different queries? |
| 04:29:30 | 8 | A. Yes, sir. |
| 04:29:30 | 9 | Q. Okay. And were you the person that was responsible? |
| 04:29:36 | 10 | A. Some of these I did directly myself, some were done with |
| 04:29:40 | 11 | my team. |
| 04:29:40 | 12 | Q. How many did you directly do by yourself? |
| 04:29:44 | 13 | A. I don't recall. |
| 04:29:44 | 14 | Q. Well, you did hundreds with the team, correct? |
| 04:29:48 | 15 | A. My expert report was 4,000 pages long, Mr. Mogin. I don't |
| 04:29:50 | 16 | recall how many I did myself. |
| 04:29:52 | 17 | Q. Well, how about in terms of percentages? |
| 04:29:58 | 18 | A. 30 percent. |
| 04:29:58 | 19 | Q. 30 percent of the Boolean queries you developed and ran? |
| 04:30:02 | 20 | A. Yes, sir. |
| 04:30:04 | 21 | Q. Okay. Now, in that case, did you use supervised learning? |
| 04:30:08 | 22 | A. I did not. |
| 04:30:08 | 23 | Q. Did you engage in statistically-ranked retrieval? |
| 04:30:12 | 24 | A. No, sir. |
| 04:30:18 | 25 | Q. Did you do statistical validation, such as has been |

04:30:22    1    discussed here today?

04:30:22    2    A.  No, sir.

04:30:24    3    Q.  Okay.  What's next in your testifying experience?

04:30:28    4    A.  I don't see or recall any other testifying experience on

04:31:12    5    this specific issue of Boolean search terms or the other areas

04:31:18    6    that you have asked for.

04:31:18    7    Q.  Okay.  So your personal hands-on experience, if you will,

04:31:26    8    please, with supervised learning, what is that?

04:31:28    9    A.  My experience with supervised learning is two-fold, but

04:31:40   10    it's not reflected in my resume.  The first is --

04:31:44   11    Q.  Let me just -- before you start so that I don't interrupt

04:31:48   12    that answer, was this resume created for this particular case;

04:31:54   13    in other words, modified or amended for this particular case?

04:31:56   14    A.  No, sir.

04:31:58   15    Q.  Did you review it before it was submitted in this

04:32:02   16    particular case?

04:32:02   17    A.  Yes, sir.

04:32:02   18    Q.  And you understood the nature of the testimony that you

04:32:06   19    would be rendering in this particular case at the time that

04:32:10   20    you submitted the resume, correct?

04:32:10   21    A.  Yes, sir.

04:32:12   22    Q.  Okay.  So please tell us about what's not on the resume

04:32:16   23    that has to do with supervised learning and your personal

04:32:20   24    hands-on experience.

04:32:22   25    A.  Yes, sir.  I have been involved in the development of

04:32:30   1   neural network programming, which is a form of supervised

04:32:32   2   learning.  Neural network programming.

04:32:38   3   Q.  And are you aware of anybody that's currently using neural

04:32:44   4   network programming in e-discovery?

04:32:46   5   A.  I'm not familiar with all the techniques people are using

04:32:54   6   in e-discovery, though network programming is widely used

04:33:02   7   globally.

04:33:02   8   Q.  For what?

04:33:02   9   A.  One example is for the detection of airplanes for

04:33:06  10   ground-based airplane safety features.  Another is for traffic

04:33:12  11   control within cities.  Another is for predicting purchaser

04:33:18  12   decisions in various industries.

04:33:20  13   Q.  So with respect to the issues that you understand you're

04:33:24  14   testifying about today, does neural network programming have

04:33:28  15   anything to do with it?

04:33:30  16   A.  It does to the extent that it is a machine-learning

04:33:34  17   technique for supervised learning.

04:33:36  18   Q.  Okay.  Does it learn from a seed set of documents?

04:33:44  19   A.  Yes.  Not documents but decisions.

04:33:46  20   Q.  So the answer then really is no, it doesn't learn from a

04:33:52  21   seed set of documents; is that right?

04:33:54  22   A.  It learns from inputs, the same as machine learning

04:34:00  23   techniques that are not neural networking learning.  You feed

04:34:06  24   it data, it develops relationships between the data you feed

04:34:10  25   it, you give it designations in the training set, later you

04:34:12  1   feed it new information from new documents or new data, and

04:34:18  2   it, based on what it's learned, makes decisions on what you

04:34:22  3   are now feeding it.

04:34:22  4   Q.  Okay.

04:34:26  5   A.  I think there is a gap here.  There's been no testimony as

04:34:32  6   to how machine-learning systems really work.

04:34:34  7   Q.  We'll get to it.

04:34:36  8   A.  I was hoping that your expert would have already covered

04:34:40  9   that.

04:34:40  10  Q.  He pretty well did cover it, Mr. Regard, but if you want

04:34:44  11  to fence, we can fence.

04:34:46  12  A.  Okay.

04:34:46  13  Q.  So you were telling us about neural network systems, and

04:34:52  14  then was there anything else that's not here on your resume

04:34:56  15  that's relevant to your testimony here today about Boolean

04:35:00  16  systems, about supervised learning, about statistically ranked

04:35:08  17  retrieval or about statistical validation in e-discovery?

04:35:12  18  A.  There are products I have worked on that are not on my

04:35:16  19  resume involving the statistical development, the development

04:35:20  20  of search terms, the improvement of the search terms, the

04:35:24  21  statistical testing of null sets, the development of query

04:35:28  22  sets, and, yes, along the development of machine learning

04:35:34  23  techniques.

04:35:34  24  Q.  And is there a reason that you didn't include those in the

04:35:40  25  material that you submitted to the court and to the parties to

04:35:44 1 establish your credentials to testify here?

04:35:46 2 A. One, when I first was engaged in this case for this

04:35:54 3 testimony, the issue before me was whether or not the

04:35:58 4 methodologies used by the defendants in preparing their search

04:36:04 5 terms for the production of documents was sufficient. And I

04:36:12 6 felt then and I still feel now that it was sufficient,

04:36:14 7 exceedingly sufficient.

04:36:16 8 Q. Based upon your understanding of best practices, right?

04:36:20 9 A. Both best practices and my personal experience in applying

04:36:24 10 those techniques within cases.

04:36:26 11 Q. You know, in your consulting practice, do you ever look at

04:36:28 12 Federal Rules of Civil Procedure?

04:36:30 13 A. Not as part of my consulting practice. As part of my

04:36:34 14 academic work and part of my work at the Sedona Conference, I

04:36:38 15 do.

04:36:40 16 Q. So you're familiar with the contours of the Federal Rules

04:36:44 17 of Civil Procedure 26?

04:36:44 18 A. Generally, yes.

04:36:46 19 Q. And how about Rule 34?

04:36:48 20 A. As well, yes.

04:36:50 21 Q. Okay. And can you tell me the best practices with respect

04:36:56 22 to ESI or search technology or any of the subjects that you

04:37:00 23 have talked about today that are in either of those rules?

04:37:06 24 A. I don't believe the rules reflect best practices in their

04:37:08 25 text.

04:37:10　1　Q.　So in your opinion, what's the relationship between best

04:37:12　2　practices and the Federal Rules?

04:37:16　3　A.　Well, first of all, not all the cases I work on are under

04:37:22　4　the Federal Rules of Civil Procedure.　I work with cases in

04:37:24　5　state court as well.

04:37:26　6　　　　　Best practices, in my opinion, is how people

04:37:28　7　implement their processes and methodology to meet their

04:37:34　8　obligations under those rules.

04:37:34　9　Q.　And in order to understand best practices, then you have

04:37:38　10　to have a pretty good understanding of the obligations,

04:37:40　11　correct?

04:37:40　12　A.　Or a lot of experience with people who are working to meet

04:37:44　13　their own obligations.

04:37:46　14　Q.　I'm sorry.　I didn't understand that.　Could you explain,

04:37:50　15　please?

04:37:50　16　A.　Certainly.　One is working with clients and comparing

04:37:56　17　notes with other thought leaders on how to achieve goals

04:38:02　18　within the litigation context, combined with the experience of

04:38:06　19　working on multiple projects over the course of many years.

04:38:14　20　Q.　Do you review the case law?

04:38:18　21　A.　Yes, sir.

04:38:18　22　Q.　Do you consider yourself current on the case law?

04:38:22　23　A.　Yes, sir.

04:38:24　24　Q.　Are you familiar with the -- let me find this case.

04:38:36　25　　　　　THE COURT:　Unless you are going to start giving

| | | |
|---|---|---|
| 04:38:38 | 1 | Judge Nolan cases.  Are you going to quiz him on something |
| 04:38:42 | 2 | else? |
| 04:38:46 | 3 | BY MR. MOGIN: |
| 04:38:48 | 4 | Q.  Are you familiar with the O'Keefe case by -- |
| 04:38:52 | 5 | THE COURT:  Judge Facciola. |
| 04:38:56 | 6 | MR. MOGIN:  Thank you, your Honor. |
| 04:38:58 | 7 | THE COURT:  Do you want the cite? |
| 04:39:02 | 8 | BY MR. MOGIN: |
| 04:39:04 | 9 | Q.  Are you familiar with that, Mr. Regard? |
| 04:39:06 | 10 | A.  I am. |
| 04:39:06 | 11 | Q.  Can you tell us what Judge Facciola has to say about key |
| 04:39:12 | 12 | words? |
| 04:39:12 | 13 | A.  I don't recall the exact -- |
| 04:39:14 | 14 | THE COURT:  I am not going to allow this unless you |
| 04:39:16 | 15 | have everything Judge Facciola has ever said on key words.  We |
| 04:39:20 | 16 | are not going to take one statement out of O'Keefe when Judge |
| 04:39:26 | 17 | Facciola has probably done more to help key word search than |
| 04:39:30 | 18 | anybody else.  And if I could go upstairs and key word John |
| 04:39:36 | 19 | Facciola and we got everything he said, I mean, there is a |
| 04:39:40 | 20 | statement in there about eeehh (phonetic), that's one |
| 04:39:48 | 21 | statement. |
| 04:39:48 | 22 | MR. MOGIN:  I am just concerned with one statement, |
| 04:39:50 | 23 | your Honor. |
| 04:39:50 | 24 | MR. NEUWIRTH:  Your Honor -- |
| 04:39:50 | 25 | THE COURT:  Okay, read it to him. |

04:39:52   1       MR. NEUWIRTH: Your Honor, this has been briefed

04:39:54   2 before Judge Shadur, and you are correct that at least there

04:39:58   3 is a strong argument to be made that this one sentence we are

04:40:02   4 about to hear is not only out of context with the rest of the

04:40:04   5 case but also with other things Judge Facciola has said. And

04:40:08   6 we would -- I think all defendants would object to the idea

04:40:12   7 that an expert is being asked --

04:40:14   8       THE COURT: Well, this is kind of contrary to what

04:40:18   9 you said to me before. Whether you think I am or not,

04:40:22  10 Mr. Mogin, you said to me, They're not supposed to be giving

04:40:26  11 legal opinions here. Now we're quoting on cases and then you

04:40:32  12 want to ask him what he thinks of what Judge Facciola said?

04:40:38  13 Is that what you want to do?

04:40:38  14       MR. MOGIN: No.

04:40:40  15       THE COURT: What do you want to do?

04:40:40  16       MR. MOGIN: Your Honor, I am not going to ask him

04:40:42  17 about legal opinions. Mr. Regard testified that he keeps up

04:40:46  18 with the case law in this area --

04:40:48  19       THE COURT: Right.

04:40:48  20       MR. MOGIN: -- that forms his position with respect

04:40:52  21 to best practices. Obviously, based upon the way that your

04:40:56  22 Honor can basically give us chapter and verse on this case,

04:40:58  23 this is an important case in the e-discovery context.

04:41:04  24       THE COURT: It is. Okay. It is.

04:41:04  25       MR. MOGIN: And so it's fair game with respect to

04:41:08  1  Mr. Regard's knowledge of this particular case as it relates

04:41:14  2  to the issues that we are dealing with today.

04:41:18  3      MR. MAROVITZ:  Judge, if I can be heard.  I was

04:41:22  4  precluded in a pretty substantial way from questioning

04:41:24  5  Mr. Regard both today and on February 21st about matters that

04:41:28  6  related to the law.  I didn't agree that I should be precluded

04:41:34  7  with respect to those rulings.  I think it is unfair now for

04:41:36  8  the plaintiff to be able to retread that same ground.

04:41:40  9      I know that -- that's in addition to the statement

04:41:42  10  that Mr. Neuwirth made.  And if, in fact, it is fair game,

04:41:46  11  then I should be permitted to ask the questions that I was not

04:41:50  12  permitted to ask before.

04:41:52  13      MR. MOGIN:  It goes to best practices, your Honor.

04:41:54  14      MR. NEUWIRTH:  But, your Honor, again, we would

04:41:58  15  object, and we can show you what we wrote to the court.  We

04:42:00  16  have formally told, I believe it was in our February 6th

04:42:04  17  brief, we believe that the out-of-context statement which I am

04:42:08  18  sure Mr. Mogin is now going to read is an incomplete and

04:42:14  19  misleading characterization of what Judge Facciola said in

04:42:16  20  this case.

04:42:18  21      THE COURT:  Well, he may have even said it in the

04:42:20  22  case, but he has a whole body of case law.

04:42:24  23      MR. NEUWIRTH:  Exactly, your Honor.

04:42:24  24      THE COURT:  He and Judge Grimm and Judge Peck have

04:42:30  25  been trying get people to cooperate, and in their quest to get

04:42:36  1  people to cooperate on an agreement on search terms, and

04:42:42  2  without pulling out their hair, they --

04:42:48  3         MR. NEUWIRTH:  And let's -- and as you know, your

04:42:50  4  Honor, what Judge Facciola was talking about in this case was

04:42:52  5  not that search terms are inaccurate; he was saying that the

04:42:56  6  defendants there were doing something wrong by suggesting that

04:43:02  7  the search terms -- that the government was using were

04:43:04  8  inadequate without making a sufficient showing of that

04:43:08  9  inadequacy.

04:43:08  10        What this quote is really saying is that arguably

04:43:12  11  what the plaintiffs are trying to do here is wrong.

04:43:14  12        MR. MOGIN:  What quote are we talking about because I

04:43:16  13  haven't asked a question yet.

04:43:18  14        THE COURT:  Okay.  Is this how you wanted to spend

04:43:22  15  the next 20 minutes or the next 25 minutes?  Is this really it

04:43:32  16  on Mr. Regard?  Do you want to ask him anything about the

04:43:34  17  chart?

04:43:36  18        MR. MOGIN:  We are going to get to the chart, your

04:43:38  19  Honor.

04:43:38  20        THE COURT:  I am telling you my chief said I have to

04:43:40  21  be someplace at 4:30.

04:43:42  22        MR. MOGIN:  I understand that.

04:43:44  23        THE COURT:  So I have to be somewhere, and I want a

04:43:46  24  few minutes to talk to the lawyers.

04:43:46  25        MR. MOGIN:  I did not understand that I was required

04:43:48  1  to finish with Mr. Regard today.  The defendants have spent

04:43:54  2  four hours cross-examining Dr. Lewis.  Time was supposed to be

04:44:00  3  equally divided.

04:44:00  4       THE COURT:  How about if you did the part of your

04:44:04  5  cross -- all right.  It's your time.  Judges get in real

04:44:26  6  trouble telling lawyers how to do their cross.

04:44:32  7  BY MR. MOGIN:

04:44:32  8  Q.  Mr. Regard, we have been talking about the O'Keefe case.

04:44:34  9  You're familiar with the case?

04:44:36  10  A.  I am.

04:44:36  11  Q.  Do you have a sense of when that case was decided?

04:44:38  12  A.  I don't recall exactly when it was decided.  The last

04:44:44  13  couple of years.

04:44:44  14  Q.  Well, if I said it was 2008, would that refresh your

04:44:46  15  recollection?

04:44:46  16  A.  It doesn't refresh --

04:44:50  17       THE COURT:  Do you agree with him?

04:44:50  18       THE WITNESS:  It doesn't refresh my recollection, but

04:44:52  19  I don't disagree.

04:44:54  20  BY MR. MOGIN:

04:44:54  21  Q.  So we can agree that I can use 2008 in my question?

04:44:56  22  A.  That's fine.

04:44:58  23  Q.  And you are aware, are you not, that in that case, the

04:45:00  24  court refers to the necessity to look to the learning that can

04:45:04  25  be provided by linguists and computer scientists, correct,

04:45:08   1   with respect to Boolean searches?

04:45:10   2   A.  That's part of what the judge references.  Can you show me

04:45:14   3   the exact text?

04:45:16   4   Q.  Okay.  Let's mark for just identification purposes -- I

04:45:22   5   guess this is Plaintiffs' 12, a copy of the decision.

04:45:32   6          MR. NEUWIRTH:  Can I have a continuing objection?

04:45:40   7          THE COURT:  Yes.

04:45:44   8   BY MR. MOGIN:

04:45:48   9   Q.  I am just going to refer you to the first page.  Do you

04:45:52  10   see the date there, February 18th, 2008?

04:45:54  11   A.  I do.

04:45:56  12   Q.  All right.  And the quote is on the last page, the second

04:46:08  13   sentence.

04:46:12  14          MR. MAROVITZ:  We really do need a copy of the

04:46:14  15   document, Judge.  Thank you.

04:46:16  16   BY MR. MOGIN:

04:46:18  17   Q.  Do you see that sentence there, please?

04:46:20  18   A.  I am on page -- I am on the last page, page 9.  Where

04:46:26  19   would you like me to look?

04:46:28  20   Q.  Yes, the last page, the second sentence.

04:46:40  21   A.  Yes, sir.

04:46:40  22   Q.  Okay.  Do you see the reference there to the knowledge of

04:46:46  23   linguists and computer scientists in connection with key word

04:46:50  24   searching?

04:46:52  25   A.  I do.

04:46:52  1  Q.  Okay.  And when did you first become familiar with this

04:46:56  2  case?  Was it shortly after it was decided?

04:46:58  3  A.  Most likely it would have been shortly after it was

04:47:04  4  decided.

04:47:04  5  Q.  So about four years ago, right?

04:47:06  6  A.  Yes, sir.

04:47:06  7  Q.  Okay.  So in those four years in connection with any of

04:47:12  8  your Boolean searches that you have been involved in, have you

04:47:16  9  engaged the services of a linguist?

04:47:20  10  A.  I have not.

04:47:20  11  Q.  Have you ever engaged the services of a linguist in

04:47:24  12  connection with the construction of Boolean key word strings?

04:47:28  13  A.  I have not engaged the services of a linguist, and in the

04:47:34  14  projects I have worked on, my clients have not felt that a

04:47:38  15  linguist would have contributed to the result.

04:47:40  16  Q.  And you yourself are not a linguist, correct?

04:47:42  17  A.  Well, I certainly understand language, and I speak several

04:47:46  18  languages, but I am not a linguist by profession.

04:47:48  19  Q.  Okay.  And let's talk about statistics.  Are you a

04:47:58  20  statistician?

04:47:58  21  A.  I do not have a doctorate degree in statistics.  I am not

04:48:04  22  sure what Mr. Lewis' degree is in.  I think it's computer

04:48:08  23  science.

04:48:08  24       I have studied statistics.  I have a degree in

04:48:12  25  computer science and mathematics, and part of my graduate

| | | |
|---|---|---|
| 04:48:16 | 1 | program included the study of statistics.  So I am familiar |
| 04:48:20 | 2 | with statistics. |
| 04:48:20 | 3 | Q.  When did you get that bachelor's degree? |
| 04:48:22 | 4 | A.  My bachelor's degree? |
| 04:48:24 | 5 | Q.  Yes, sir. |
| 04:48:26 | 6 | A.  1989. |
| 04:48:30 | 7 | Q.  And when did you get the MBA? |
| 04:48:30 | 8 | A.  1995. |
| 04:48:32 | 9 | Q.  And have you taken any classes in statistics at the |
| 04:48:36 | 10 | graduate level since 1995? |
| 04:48:38 | 11 | A.  No, sir. |
| 04:48:40 | 12 | Q.  Have you written any papers on statistics since 1995 that |
| 04:48:48 | 13 | have been peer reviewed? |
| 04:48:48 | 14 | A.  No, sir. |
| 04:48:50 | 15 | Q.  Have you taught statistics since 1995 in a college class |
| 04:48:56 | 16 | or university class? |
| 04:48:56 | 17 | A.  Not in a university class or college class, I have not.  I |
| 04:49:08 | 18 | was asked to teach statistics to Georgetown advanced institute |
| 04:49:14 | 19 | for e-discovery last year, and I did conduct a session on |
| 04:49:18 | 20 | statistical use within e-discovery. |
| 04:49:18 | 21 | Q.  Well, you were one of how many panelists? |
| 04:49:22 | 22 | A.  I was one of three panelists and one moderator. |
| 04:49:28 | 23 | Q.  And did you -- |
| 04:49:28 | 24 | A.  Three panelists, I think so, three. |
| 04:49:32 | 25 | Q.  Did you present a paper that you authored? |

04:49:34  1  A.  No, sir.

04:50:16  2        MR. MOGIN:  Your Honor, I'm going to mark as next in

04:50:18  3  order Temple-Inland's responses and objections to plaintiffs'

04:50:24  4  first request for production of documents directed to all

04:50:26  5  defendants.

04:50:26  6        THE COURT:  What number would that be?

04:50:28  7        MR. MOGIN:  I believe that's 13, Plaintiffs' 13.

04:50:30  8        THE COURT:  This is Plaintiffs' 13?

04:50:32  9        MR. NEUWIRTH:  Yes.

04:50:32  10        THE COURT:  Thank you.

04:50:42  11  BY MR. MOGIN:

04:50:42  12  Q.  Mr. Regard, are you familiar with Plaintiffs' 13?

04:50:46  13  A.  I don't recall sitting here if that was one of the

04:50:50  14  documents I reviewed or not.  Many, if not most, of these were

04:51:02  15  produced to me.

04:51:02  16  Q.  Do you recall if this one was?

04:51:06  17  A.  I don't offhand.

04:51:08  18  Q.  Do you want to look at Exhibit B of that report?

04:51:18  19        THE COURT:  What page are you on?

04:51:22  20        MR. MOGIN:  It's page 1 of Exhibit B.

04:51:24  21        THE COURT:  Page 1 of Exhibit B.

04:51:38  22        THE WITNESS:  I have listed all the materials that I

04:51:40  23  considered.  I am sure I looked at it at some point, I just

04:51:46  24  don't recall from looking on the face of it.

04:51:48  25  BY MR. MOGIN:

| | | |
|---|---|---|
| 04:51:48 | 1 | Q.  Can you identify for us any document request that |
| 04:51:54 | 2 | Temple-Inland did not tender an objection for? |
| 04:51:58 | 3 | A.  I haven't read the document at that level of degree. |
| 04:52:06 | 4 | Would you like for me to? |
| 04:52:06 | 5 | THE COURT:  No. |
| 04:52:10 | 6 | Sorry. |
| 04:52:12 | 7 | BY MR. MOGIN: |
| 04:52:12 | 8 | Q.  Will you accept my representation that there are none? |
| 04:52:14 | 9 | A.  Are you suggesting they are not responding to any of your |
| 04:52:20 | 10 | requests or they just lodged and preserved their objections to |
| 04:52:24 | 11 | your requests? |
| 04:52:26 | 12 | Q.  The latter. |
| 04:52:28 | 13 | A.  I would accept your representation. |
| 04:52:30 | 14 | Q.  Now, are you aware of which document requests the search |
| 04:52:38 | 15 | strings refer to? |
| 04:52:40 | 16 | A.  No, sir. |
| 04:52:42 | 17 | Q.  Do you have any idea how many of the document requests are |
| 04:52:50 | 18 | within the search strings? |
| 04:52:50 | 19 | A.  I believe there's documentation to the effect of which |
| 04:52:56 | 20 | document requests Temple has agreed to respond to.  I don't |
| 04:53:02 | 21 | have -- I don't know the correlation of which ones of those |
| 04:53:06 | 22 | were effected through the use of search strings versus which |
| 04:53:08 | 23 | ones were effected through other document collections, review, |
| 04:53:12 | 24 | and productions. |
| 04:53:14 | 25 | Q.  So are you suggesting that there's documentation about |

04:53:16  1  requests that Temple has agreed to other than their formal

04:53:20  2  response?

04:53:20  3  A.  I am trying to remember if I have seen documentation

04:53:34  4  reflecting search strings against individual requests.  I

04:53:38  5  don't recall sitting here.  If I did, it was through the

04:53:42  6  documentation that was produced by defendants to plaintiffs.

04:53:44  7  Q.  So in your consideration of your opinions in this case,

04:53:52  8  did you consider whether or not the defendants had attempted

04:53:56  9  to construct search strings -- strike that.

04:54:00  10        Were you aware that in approximately July of 2011 at

04:54:08  11  the defendants' request, the plaintiffs broke the request for

04:54:14  12  production of document into several categories, one was

04:54:18  13  conduct requests, another was data requests, the third had to

04:54:22  14  do with inquiries and investigations, and the fourth was more

04:54:28  15  general?

04:54:28  16  A.  I don't recall that specific correspondence, no.

04:54:30  17  Q.  In fact, is that correspondence listed anywhere in your

04:54:40  18  documentation that you looked at?

04:54:42  19  A.  I don't see it offhand.

04:55:02  20  Q.  Okay.

04:55:02  21  A.  It may -- many of the earlier correspondence are, in fact,

04:55:10  22  attachments to some of these correspondences, so I can't say

04:55:14  23  definitively that I did not receive it.  I don't recall it.

04:55:16  24  Q.  All right.  Regardless of whether you received it or

04:55:22  25  reviewed it, were you aware that this had occurred?

04:55:26 | 1 | A.  I am now.

04:55:28 | 2 | Q.  Prior to my bringing it to your attention, were you aware

04:55:32 | 3 | of it?

04:55:32 | 4 | A.  I think I was aware that there had been a distinction made

04:55:36 | 5 | between data requests, meaning non-emails and non-user files,

04:55:42 | 6 | versus the requests that the defendants were seeking to

04:55:46 | 7 | satisfy through the collection and review of documents, but I

04:55:50 | 8 | don't recall sitting here seeing that categorized in the way

04:55:56 | 9 | that you have described it.

04:55:56 | 10 | Q.  Would it surprise you to learn -- strike that.

04:56:02 | 11 | Have you reviewed the documentation that

04:56:06 | 12 | Georgia-Pacific tendered last week having to do with search

04:56:10 | 13 | string validation?

04:56:12 | 14 | A.  I am.

04:56:12 | 15 | Q.  Would it surprise you to learn that the defendants

04:56:18 | 16 | responded or tried to create search strings to only

04:56:22 | 17 | approximately 17 document requests?

04:56:24 | 18 | A.  I believe you just told me that they objected to all of

04:56:32 | 19 | your document requests --

04:56:32 | 20 | Q.  That's correct.

04:56:34 | 21 | A.  -- so that does not surprise me.

04:56:36 | 22 | Q.  Okay.  And did you in any way consider that in forming

04:56:40 | 23 | your opinions today as to whether or not defendants were

04:56:46 | 24 | compliant with best practices?

04:56:48 | 25 | A.  I don't find that relevant to the consideration of their

04:56:52   1   compliance with best practices.

04:56:54   2   Q.  Did you consider today, in connection with forming your

04:56:58   3   opinions that you have talked about today, did you consider

04:57:02   4   the fact that prior to receiving the documents last week from

04:57:08   5   Georgia-Pacific, the defendants had never specifically

04:57:12   6   identified which particular document request the search

04:57:18   7   strings were intended to respond to?

04:57:20   8   A.  I did not consider that, no, sir.

04:57:24   9   Q.  Do you consider such conduct to be within the realm of

04:57:28   10  transparency as you described it earlier?

04:57:34   11  A.  Back to your former question, you asked me if I considered

04:57:38   12  whether or not the defendants had considered which document

04:57:42   13  requests in responding to the development of search terms.  I

04:57:48   14  think my earlier testimony was my information is such that

04:57:50   15  they considered your request for documents very heavily in

04:57:54   16  developing their search terms.  That was part of the input

04:58:00   17  that went into the front of the process.

04:58:02   18        Exactly which ones they decided to respond to versus

04:58:04   19  not is definitely an important aspect that they made that

04:58:08   20  decision, not so much that I know which ones they decided to

04:58:12   21  respond to and which ones they did not decide to respond to.

04:58:16   22  Q.  Okay.  So 17 out of 94, does that strike you as consistent

04:58:22   23  with best practices?

04:58:24   24  A.  It strikes me as irrelevant to best practices.

04:58:26   25  Q.  Okay.  And does the fact that plaintiffs were not advised

04:58:32   1   which of the 94 the search strings related to prior to these

04:58:38   2   hearings have any bearing on your testimony with respect to

04:58:42   3   the transparency of the process?

04:58:44   4   A.  It does and it doesn't at the same time.  It does in that

04:59:02   5   I believe that describing the process that went through to

04:59:08   6   derive search terms, apply the search terms, and test the

04:59:12   7   results, that process I feel has had transparency.  I also was

04:59:20   8   informed when the search terms were produced to the

04:59:24   9   plaintiffs.

04:59:24  10          If there had been omissions there or that the

04:59:26  11   plaintiffs had felt, and I think you did in occasions, that

04:59:30  12   the search terms would not have satisfied your document

04:59:32  13   requests, you advised them, and they took that into

04:59:34  14   consideration in the revision of their search terms.

04:59:40  15          As to the communication of the individual search

04:59:42  16   terms that are being requested or being responded to versus

04:59:48  17   not responded to, I think this document is pretty clear, as

04:59:52  18   you have represented it, that they're objecting to all of your

04:59:54  19   requests.

04:59:58  20   Q.  If they didn't tell us what the search strings

05:00:02  21   corresponded to in terms of the particular requests for

05:00:06  22   production of documents, do you consider that to be

05:00:08  23   transparent in the meaning of transparency as you earlier

05:00:14  24   testified about in connection with best practices?

05:00:16  25   A.  I think it would be important for me to understand the

| 05:00:22 | 1 | chronology of events more clearly than you have laid them out |
| 05:00:26 | 2 | here. I don't know when those decisions were made, and I |
| 05:00:32 | 3 | don't know how that correlated to decisions of plaintiffs. |
| 05:00:36 | 4 | Q. You are suggesting then that I have to walk you through |
| 05:00:38 | 5 | each of the search string developments -- |
| 05:00:40 | 6 | THE COURT: What number are your search strings? |
| 05:00:44 | 7 | MR. MOGIN: The search strings are -- |
| 05:00:48 | 8 | THE COURT: No. They are in the requests? |
| 05:00:50 | 9 | MR. MOGIN: No, no. |
| 05:00:56 | 10 | Your Honor, just to be clear, we didn't give them |
| 05:00:58 | 11 | search strings. They made them up. |
| 05:01:00 | 12 | THE COURT: So that chart with the highlighting, the |
| 05:01:04 | 13 | one you dropped off? |
| 05:01:06 | 14 | MR. MOGIN: That goes with Ms. Tenny. It's a |
| 05:01:08 | 15 | demonstrative. |
| 05:01:10 | 16 | THE COURT: That's Ms. Tenny. Okay. |
| 05:01:12 | 17 | MR. MOGIN: She is going to kill me. |
| 05:01:34 | 18 | BY MR. MOGIN: |
| 05:01:34 | 19 | Q. You are suggesting then I am going to have to walk you |
| 05:01:38 | 20 | through all of the correspondence and all of the search string |
| 05:01:42 | 21 | development documents that we received last week before you |
| 05:01:44 | 22 | can answer that question? |
| 05:01:44 | 23 | THE COURT: Or maybe if one of the defendants know |
| 05:01:48 | 24 | the answer, they could stipulate to whether or not -- say the |
| 05:01:54 | 25 | question again, please. I was busy reading the responses. I |

05:01:58    1    thought we were still on the first request for production.  I

05:02:02    2    was reading those.

05:02:04    3    BY MR. MOGIN:

05:02:04    4    Q.  Are you suggesting, Mr. Regard, that in order for you to

05:02:08    5    be able to answer my question, that we have to trace through

05:02:12    6    all of the correspondence and all of the search string

05:02:16    7    developments so that you can tell us about the transparency

05:02:20    8    that took place or didn't take place and whether that's

05:02:26    9    relevant to your opinion?

05:02:26    10    A.  The transparency that I testified to earlier is the

05:02:30    11    transparency of Boolean search terms, that when they are

05:02:34    12    produced, they can be read and they can be restored.  To the

05:02:36    13    extent they were provided to you, the plaintiffs, I have seen

05:02:40    14    correspondence that you responded to them, so the transparency

05:02:46    15    characteristic of search terms to be readable by people

05:02:50    16    obviously had an effect because you responded to them.

05:02:54    17            Separately, I have testified that it's important for

05:02:56    18    parties to relate to the counter-party the process they're

05:03:02    19    going through.  I'm not trying to muddle those two concepts

05:03:06    20    together.

05:03:08    21    Q.  Well, should the process be related to the counter-party

05:03:10    22    contemporaneously or after the fact?

05:03:14    23    A.  That varies from party to party, from litigation to

05:03:22    24    litigation.

05:03:22    25    Q.  What's best practice?

05:03:26  1  A.  I think that still relates from party to party to

05:03:28  2  litigation to litigation.

05:03:28  3  Q.  What's best practice under the Sedona principles?

05:03:32  4  A.  Under Sedona, it's to balance that transparency with your

05:03:38  5  client's interest and your advocacy.

05:03:40  6  Q.  When you advise clients, do you suggest to them that they

05:03:44  7  create the equivalent of an audit trail with respect to their

05:03:48  8  e-discovery processes?

05:03:50  9  A.  I don't usually advise clients of that.  I usually take

05:03:58  10  care of that in my company when we are applying our services.

05:04:02  11  We keep documentation usually on cases we work on and the

05:04:08  12  steps we take.

05:04:08  13  Q.  Was that done in this case?

05:04:10  14  A.  I didn't develop the search terms in this case.

05:04:12  15  Q.  But you did testify that the defendants complied in your

05:04:14  16  opinion with best practices?

05:04:18  17  A.  That's correct.

05:04:18  18  Q.  Okay.  So does best practices require some level of

05:04:24  19  replicability, accountability, the ability to recreate the

05:04:32  20  investigation?

05:04:32  21  A.  In best practices, yes, you'd want to have some

05:04:38  22  documentation of what you have done and the ability to

05:04:40  23  validate each individual step.

05:04:42  24  Q.  So when Mr. Brown testified that after he, quote,

05:04:48  25  rationalized the search strings, he was no longer able to

| | | |
|---|---|---|
| 05:04:54 | 1 | correlate any particular search string to any particular |
| 05:04:56 | 2 | request, was that, in your opinion, consistent with best |
| 05:04:58 | 3 | practices as you've just described the need for replicability, |
| 05:05:04 | 4 | transparency? |
| 05:05:04 | 5 | THE COURT:  We have an objection. |
| 05:05:06 | 6 | MR. NEUWIRTH:  Lack of foundation.  That's not what |
| 05:05:08 | 7 | Mr. Brown said, and I have his testimony on that point right |
| 05:05:10 | 8 | in front of me. |
| 05:05:12 | 9 | THE COURT:  Well, you have to be accurately stating |
| 05:05:14 | 10 | what the testimony was, so take a look at it. |
| 05:06:08 | 11 | MR. MOGIN:  People are searching for it. |
| 05:06:08 | 12 | THE COURT:  You know what?  It's five after 4:00.  I |
| 05:06:12 | 13 | do actually want to talk to the lawyers.  This is a very hard |
| 05:06:14 | 14 | group to get together.  Folks who are not the lawyers, we |
| 05:06:18 | 15 | don't have enough time to go down to my chambers at the |
| 05:06:22 | 16 | moment, so I would just like to talk to you here in the |
| 05:06:24 | 17 | courtroom.  It's, of course, a public courtroom, but I would |
| 05:06:28 | 18 | really appreciate it if the visitors would leave and I could |
| 05:06:30 | 19 | just talk to the lawyers here. |
| 05:06:32 | 20 | I don't know -- it somewhat -- what I have to say |
| 05:06:36 | 21 | somewhat involves Dr. Lewis and Dan Regard, so if you want |
| 05:06:40 | 22 | them to stay, it's fine with me. |
| 05:06:44 | 23 | MR. MAROVITZ:  That's fine. |
| 05:06:44 | 24 | THE COURT:  Okay? |
| 05:06:46 | 25 | Dr. Tenny, where are you?  Thank you so much.  You |

| | | |
|---|---|---|
| 05:06:50 | 1 | are the best sport for coming two days.  We will be back to |
| 05:07:00 | 2 | you shortly on what our next step will be here. |
| 05:07:04 | 3 | You can step down, Mr. Regard. |
| 05:07:06 | 4 | THE WITNESS:  Thank you, your Honor. |
| 05:07:06 | 5 | (Witness leaves the stand.) |
| 05:07:06 | 6 | THE COURT:  We are going to be finished if you are |
| 05:07:08 | 7 | running for a plane.  We are going to be finished in a few |
| 05:07:10 | 8 | minutes, I promise you. |
| 05:08:18 | 9 | Okay.  So this is what I am calling, if everybody can |
| 05:08:20 | 10 | hear me, when I received this referral in December, it was |
| 05:08:30 | 11 | from my darling Judge Shadur, who never sends anything to me. |
| 05:08:34 | 12 | I am putting this in context.  Okay?  And it said, Conduct an |
| 05:08:38 | 13 | evidentiary hearing.  So I thought -- being the obedient |
| 05:08:44 | 14 | little magistrate judge that I am, I thought, Oh, I better set |
| 05:08:48 | 15 | this thing real fast.  So we did.  And this is quite a large |
| 05:08:54 | 16 | group here. |
| 05:08:56 | 17 | If I had one thing to do over, and that's what I am |
| 05:09:00 | 18 | doing right now, if we could go back to that day, and I had |
| 05:09:06 | 19 | reviewed as many documents as I now have reviewed, I want to |
| 05:09:12 | 20 | say a couple things because I don't think it's too late to go |
| 05:09:16 | 21 | back to that original place. |
| 05:09:18 | 22 | I am a believer of principle 6 of Sedona, and I'm not |
| 05:09:28 | 23 | just because it's Sedona, but I think the people who are |
| 05:09:30 | 24 | producing the records, producing the documents, are in a |
| 05:09:36 | 25 | better position to know, since they have to do the work, spend |

05:09:42  1   the money, spend the time, they know their people, they know

05:09:46  2   their material, so as a basic premise, I think that's a pretty

05:09:50  3   fair premise here.

05:09:52  4          I also think I don't quite understand why they went

05:10:00  5   so fast without getting you involved.  Okay?  But as soon as

05:10:06  6   they found out, I'm sort of assuming, they were trying to

05:10:10  7   figure out, all seven of them, of what the heck was going on,

05:10:14  8   I'd say by August, and Judge Shadur did not decide his motion

05:10:18  9   to dismiss until April.  So by August, I think the dialog

05:10:24  10  starts.

05:10:26  11         And the reason I'm even going to suggest what I am

05:10:30  12  going to suggest is I think -- I would give you a B plus for

05:10:38  13  cooperation, communication with each other.  I actually think

05:10:42  14  you really did -- once it got started, I think you did a

05:10:48  15  really good job.

05:10:50  16         I don't know whether the indexing issue started to

05:10:54  17  send this off on this kind of I'd almost call it a detour

05:11:02  18  we're on with quote, unquote, predictive quoting, what all the

05:11:08  19  blogs are talking about us.

05:11:10  20         I assume you and Dr. Lewis, what you really are

05:11:16  21  interested in is a search, regardless if it's Boolean or

05:11:24  22  computer-assisted, that is fair and statistically -- and that

05:11:32  23  can be validated statistically because that would be a good

05:11:38  24  word search.

05:11:40  25         My question to all of you right now -- really, it's

05:11:44　1　to you, and you don't have to answer me today; I even have a

05:11:50　2　time we are going to call up again -- is there a way, rather

05:11:56　3　than starting all over with all of the good work that is here,

05:12:04　4　if Dr. Lewis and Mr. Regard were able to help and we were able

05:12:12　5　to within a same framework take their search and be able to

05:12:20　6　tweak it and make it something that you could be comfortable

05:12:24　7　with?

05:12:24　8　　　　　MR. MOGIN:  You're right.  I won't answer you today.

05:12:26　9　　　　　THE COURT:  You don't have to answer me today.  And I

05:12:30　10　understand that I am sort of -- and I have no idea.  Now, I

05:12:36　11　have no idea if they want to go to the mat as the Godfather

05:12:40　12　would say, with their search as it is.  Maybe they don't want

05:12:44　13　to tweak anything.

05:12:46　14　　　　　MR. MOGIN:  I will --

05:12:46　15　　　　　THE COURT:  I don't know.  I don't know.

05:12:50　16　　　　　What I have gotten out of this -- and I think there

05:12:58　17　is a bigger hole in the case book of what is statistically

05:13:04　18　correct.  I have been walking around saying to Chris for the

05:13:10　19　last week, because we spend so much time fighting with parties

05:13:16　20　about agree, agree, agree, all of the case law, all of my

05:13:22　21　time, all of my opinions in this area, as Joe knows, is

05:13:26　22　beating people over the head to agree what the search terms

05:13:30　23　are.  We never get to -- I really, you know, you can all jump

05:13:34　24　in here.  You start telling me cases you know, other than Paul

05:13:40　25　Grimm kind of waxing on that it should be statistically valid,

| | | |
|---|---|---|
| 05:13:48 | 1 | and what Judge Facciola said, there hasn't -- judges just |
| 05:13:54 | 2 | haven't -- parties haven't been thinking like that, judges |
| 05:13:58 | 3 | haven't been thinking like that, and I actually think that's |
| 05:14:02 | 4 | probably the more helpful part of the case is what I am trying |
| 05:14:08 | 5 | to say.  And I happen to think it should be, it should be |
| 05:14:14 | 6 | valid.  It shouldn't be, Oh, my God, oh, my God, let's just |
| 05:14:18 | 7 | move on, let's just get rid of this. |
| 05:14:22 | 8 | So your homework assignment is I want you to talk to |
| 05:14:28 | 9 | each other, see if Dan would come up with -- if Dan would |
| 05:14:34 | 10 | agree that he would go -- he would go along with the Boolean |
| 05:14:38 | 11 | search and he tell you what kind of changes, what kind of |
| 05:14:44 | 12 | tweaking, what kind of running, whatever he needs, whatever |
| 05:14:50 | 13 | kind of validation they need, if you would be willing to do |
| 05:14:52 | 14 | that. |
| 05:14:56 | 15 | In exchange for that, here is my quid pro quo.  We |
| 05:15:00 | 16 | could take -- it is five months tomorrow that I leave here.  I |
| 05:15:06 | 17 | will work with you for the next five months on trying to |
| 05:15:08 | 18 | figure out privilege, indexing, 30(b)(6)s.  We could take the |
| 05:15:18 | 19 | five months and try to get you in some kind of shape where |
| 05:15:24 | 20 | maybe you could get your arms around the rest of the discovery |
| 05:15:28 | 21 | issues here. |
| 05:15:30 | 22 | I actually think -- I mean, when you say, you know, |
| 05:15:34 | 23 | you're not exaggerating that you could be coming in on motions |
| 05:15:38 | 24 | to compel on this for the next two years to whoever the new |
| 05:15:42 | 25 | magistrate judge is. |

05:15:46   1    So I don't know.  I think it's kind of a matter of

05:15:48   2  where you want to put your resources.  I know you all have

05:15:54   3  clients, and you've got all those other plaintiffs' lawyers,

05:15:58   4  and I mean it as somebody who used to get kicked around every

05:16:04   5  courtroom in this building as a criminal defense lawyer, you

05:16:06   6  can imagine, I am not going to take this personal, but I did

05:16:10   7  not want to walk out of here today and not say to you, Hey,

05:16:16   8  why don't we all take a nice, big, deep breath, step back, and

05:16:22   9  see if there's not something we could do to save this right

05:16:26   10  now.  And not only save it, make it better.

05:16:28   11    MR. MOGIN:  I appreciate what you're saying.  I will

05:16:30   12  give it good-faith consideration.

05:16:32   13    THE COURT:  Good.

05:16:32   14    MR. MOGIN:  However, just so defense counsel hears it

05:16:36   15  loud and clear, so that they can't accuse us of any holding

05:16:40   16  back, since December 2010 when Mr. Neuwirth and I had a heated

05:16:48   17  exchange in the hallway, we have said we will not tolerate a

05:16:52   18  search that is restricted to these custodians.  It won't -- we

05:16:58   19  will not make that agreement.

05:17:00   20    THE COURT:  So that would be one of the things you

05:17:06   21  would come back, is you would want to propose custodians?

05:17:10   22    MR. MOGIN:  No.  It's going to have to be some other

05:17:12   23  way besides custodians.

05:17:16   24    THE COURT:  Oh, you won't do a custodian-based

05:17:18   25  search.

05:17:18   1      MR. MOGIN:  I certainly won't do these top custodians

05:17:20   2  that they have picked out.

05:17:22   3      THE COURT:  I see.

05:17:24   4      MR. ECHOLS:  Your Honor, Barack Echols.  Ms. Miller

05:17:28   5  and I, during the course of the summer and after, have had

05:17:30   6  some of those conversations.  I think you saw some of that

05:17:32   7  correspondence that we sent to you.  And our position has

05:17:36   8  always been that this is our position as to the ones that make

05:17:40   9  the most sense.  We think this may be all.  We understand you

05:17:44  10  disagree, we understand there may have to be a conversation

05:17:48  11  about additional custodians at some point, but we never have

05:17:50  12  been able to get into that, even at which seems to be a

05:17:54  13  reasonable place to be because, as you said, we took a little

05:17:58  14  bit of a detour.

05:18:00  15      THE COURT:  Can you do -- experts, could you do I

05:18:06  16  would call what Mr. Mogin is saying as some kind of concept or

05:18:12  17  a broader-based search, what would you call it, other than the

05:18:16  18  non-custodial?  Can you do that with Boolean?

05:18:24  19      DR. LEWIS:  You are asking me?

05:18:24  20      THE COURT:  Yes.

05:18:26  21      DR. LEWIS:  That is can you do a Boolean search on

05:18:28  22  material that wasn't gathered by custodians but were gathered

05:18:30  23  some other way?

05:18:32  24      THE COURT:  Yes.

05:18:32  25      DR. LEWIS:  Yes.

05:18:32  1    THE COURT:  Clusters or concepts or any --

05:18:34  2    DR. LEWIS:  Well, no, I'm saying -- you are talking

05:18:36  3  about the collection procedure, whether it's collection --

05:18:38  4    THE COURT:  You're right.  But you could use Boolean

05:18:40  5  for a non-custodial search?

05:18:44  6    DR. LEWIS:  Yes.

05:18:44  7    THE COURT:  All right.

05:18:50  8    MR. McKEOWN:  The scope of what is going to be in the

05:18:52  9  universe is a very big question.  We have other custodial

05:18:56  10  documents that we have collected and are reviewing, and we

05:18:58  11  have shared servers that we have collected and are reviewing.

05:19:02  12  But if it's -- we have to take every document in the company,

05:19:04  13  and that's a major problem.

05:19:06  14    THE COURT:  When I would like to talk to you is

05:19:08  15  Tuesday at 4:00 o'clock, if I can, Chicago time.  And one of

05:19:14  16  you must have a bridge line with all your technology --

05:19:22  17    MS. MILLER:  We will provide it, your Honor.  We will

05:19:24  18  take care of it.

05:19:26  19    THE COURT:  And at least maybe you will give one

05:19:28  20  call, somebody, whoever has the most charm, some Irish person,

05:19:34  21  some other Irish person, and try to have a conversation here

05:19:40  22  before Tuesday, and if the answer is no, the answer is no.

05:19:44  23  And then we know what to do.  We will go back to -- we will

05:19:50  24  figure out who else has to be heard on the hearing.  I have

05:19:54  25  Friday, April 29th, a full day.

| | | |
|---|---|---|
| 05:20:04 | 1 | MS. MILLER: April 29th is a Sunday. |
| 05:20:06 | 2 | THE COURT: Thank you. April 27th. I do have |
| 05:20:10 | 3 | Friday, April 27th, open. We could do another round if you |
| 05:20:16 | 4 | wanted to have another day. That gives you enough time. This |
| 05:20:22 | 5 | group is just so darn hard -- now, here is the other thing. |
| 05:20:26 | 6 | That's my suggestion in here. I would love to hear anybody |
| 05:20:32 | 7 | else's suggestion, what we could do, short of starting over |
| 05:20:38 | 8 | from scratch, and that would free Chris and I up. |
| 05:20:48 | 9 | If we spend all of our time, regardless of how we |
| 05:20:52 | 10 | turn out, I am not available to you to do anything else. And |
| 05:20:58 | 11 | I am really good at some form of mediation in this. So it's |
| 05:21:02 | 12 | kind of -- so if you have any other suggestion of what we |
| 05:21:10 | 13 | could do for the next five months, I also want to hear that. |
| 05:21:18 | 14 | MR. MAROVITZ: Judge, can I -- |
| 05:21:20 | 15 | THE COURT: Sure. |
| 05:21:20 | 16 | MR. MAROVITZ: Far be it from me to throw a kink in |
| 05:21:24 | 17 | the works here, and I hope that we wouldn't need the 27th. I |
| 05:21:28 | 18 | have a multiparty -- I can move many things, I have a |
| 05:21:30 | 19 | multiparty mediation scheduled for that day. I cannot move |
| 05:21:34 | 20 | that. |
| 05:21:34 | 21 | THE COURT: So Tuesday then. When we talk Tuesday, I |
| 05:21:38 | 22 | will have to have -- you know, we might have to do it on a |
| 05:21:42 | 23 | Saturday. |
| 05:21:42 | 24 | MR. MAROVITZ: I apologize. |
| 05:21:44 | 25 | THE COURT: Maybe we will just do it on a Saturday or |

05:21:46 1 something in New York.

05:21:50 2          MR. MOGIN:  San Diego.

05:21:52 3          THE COURT:  Or San Diego.  Absolutely.

05:21:54 4          Anybody want to say anything?  You don't even have to

05:21:58 5 be on the record.

05:22:02 6          MR. McKEOWN:  We will communicate prior to Tuesday.

05:22:04 7          THE COURT:  Good.  And you talk to each other before

05:22:06 8 Tuesday and see if, you know, I am just being the old dreamer

05:22:12 9 that I am.

05:22:14 10          Thank you, everybody.

05:22:16 11          MR. McKEOWN:  Thank you.

05:22:16 12          MS. MILLER:  Thank you, your Honor.

05:22:18 13          MR. MAROVITZ:  Judge, one other final note just in

05:22:20 14 terms of my pet project, the witness rule.  Again, I am

05:22:24 15 hopeful that we won't need to come back, but if we do, can I

05:22:28 16 talk to Mr. Regard?

05:22:28 17          THE COURT:  You may talk to Mr. Regard, you may talk

05:22:30 18 to Dr. Lewis, you may talk to Ms. Tenny.  You may talk to --

05:22:38 19 Dr. Tenny, yes, excuse me.

05:22:40 20          MR. MAROVITZ:  Thank you, Judge.

05:22:40 21          THE COURT:  And you can talk to anybody else you

05:22:42 22 want, Mr. Marovitz.

05:22:44 23          Bye, everybody.

05:22:46 24          MS. MILLER:  Thank you, your Honor.

05:22:46 25          MR. McKEOWN:  Thank you, your Honor.

| | | |
|---|---|---|
| 05:22:48 | 1 | MR. NEUWIRTH:  Thank you, your Honor. |
| 05:22:56 | 2 | THE COURT:  Hold on. |
| 05:23:00 | 3 | MR. CAMPBELL:  I just think the question needs to be |
| 05:23:04 | 4 | about how do we satisfy our comfort level if the response -- a |
| 05:23:12 | 5 | vacuum is produced.  And I would hope that things don't have |
| 05:23:14 | 6 | to be taken off the table as long as you get to that comfort |
| 05:23:18 | 7 | level because that what we are talking about, producing |
| 05:23:20 | 8 | responsive documents. |
| 05:23:22 | 9 | THE COURT:  That's the way we are seeing it.  In the |
| 05:23:26 | 10 | end, that's what your responsibility to your clients are. |
| 05:23:32 | 11 | Okay?  Thanks, everybody. |
| 05:23:34 | 12 | MR. McKEOWN:  Thank you, your Honor. |
| 05:23:34 | 13 | MS. MILLER:  Thank you. |
| | 14 | (Which were all the proceedings had in the above-entitled |
| | 15 | cause on the day and date aforesaid.) |
| | 16 | I certify that the foregoing is a correct transcript from |
| | 17 | the record of proceedings in the above-entitled matter. |
| | 18 | Carolyn R. Cox                              Date |
| | | Official Court Reporter |
| | 19 | Northern District of Illinois |
| | 20 | /s/Carolyn R. Cox, CSR, RPR, CRR, FCRR |
| | 21 | |
| | 22 | |
| | 23 | |
| | 24 | |
| | 25 | |